

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:53:45 ; Search time 95.3333 Seconds  
(without alignments)  
1934.128 Million cell updates/sec

Title: US-09-987-701-4

Perfect score: 3649

Sequence: 1 MSISDEVNVLVRYLQESG.....LPHHLVIVLPIELLVLK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: A Geneseq\_23Sep04:\*

- 1: Genesecp1980s:\*
- 2: Genesecp1990s:\*
- 3: Genesecp2000s:\*
- 4: Genesecp2001s:\*
- 5: Genesecp2002s:\*
- 6: Genesecp2003as:\*
- 7: Genesecp2003bs:\*
- 8: Genesecp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	3437	94.2	514	8 ADJ75513	Adj75513 Marker ge
2	3428	93.9	514	4 AAB95225	Aab95225 Human pro
3	3427	93.9	514	6 ABO07190	Abo07190 Human p53
4	3402	93.2	514	8 ADJ76263	Adj76263 Marker ge
5	3154.2	86.4	577	7 ADD14051	Add14051 Human src
6	3154.2	86.4	577	8 ADQ18019	Adq18019 Human sof
7	3069.7	84.1	459	4 ADM19860	Adm19860 Protein e
8	3064.6	84.0	542	5 ABP41760	Abp41760 Human ova
9	2753.9	75.5	700	4 ABB60376	Abb60376 Drosophila
10	2744	75.2	395	5 ABP51424	Abp51424 Human MDD
11	2643.1	72.4	584	4 ABG21351	Abg21351 Novel hum
12	2605.2	71.4	7064	7 ADG70546	Adg70546 Aspergill
13	2601.6	71.3	5635	5 ABP60991	Abp60991 Novel hum
14	2596.5	71.2	5636	7 ADJ70089	Adj70089 Human hea
15	2596.5	71.2	5636	7 ADJ83137	Adj83137 Human hem
16	2596.5	71.2	5636	8 ADK60205	Adk60205 Angiogene
17	2596.5	71.2	5636	8 ADK60506	Adk60506 Angiogene
18	2596.5	71.2	5636	8 ADP73129	Adp73129 Angiogene
19	2591.5	71.0	31267	6 ABG74786	Abg74786 Human RGS
20	2589.5	71.0	4599	8 ADI27168	Adi27168 Mouse LRP
21	2589.5	71.0	4599	8 ADI27169	Adi27169 Mouse LRP
22	2580.2	70.7	7107	4 ABB58144	Abb58144 Drosophila
23	2576.4	70.6	5509	8 ADN96832	Adn96832 Bugula br
24	2576.2	70.6	6940	8 ADN96829	Adn96829 Bugula br
25	2575.8	70.6	4599	6 ABP56837	Abp56837 Human LRP

## ALIGNMENTS

## RESULT 1

ADJ75513  
ID ADJ75513 standard; protein; 514 AA.

XX AC ADJ75513;

XX DT 20-MAY-2004 (first entry)

XX DE Marker gene related amino acid sequence SEQ ID NO:765.

XX KW bronchial asthma; chronic obstructive pulmonary disease;

KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;

KW gene therapy; marker.

XX OS Homo sapiens.

XX PN EP1394274-A2.

XX PD 03-MAR-2004.

XX PF 04-AUG-2003; 2003EP-00254857.

XX PR 06-AUG-2002; 2002JP-00229312.

XX PR 20-MAR-2003; 2003JP-00077212.

XX PA (GENO-) GENOX RES INC.

XX PI Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;

XX DR WPI; 2004-193155/19.

XX PT Testing for bronchial asthma or chronic obstructive pulmonary disease by  
comparing the expression level of a marker gene in a biological sample  
from a subject with the expression level of the gene in a sample from a  
healthy subject.

XX PS Example 11; SEQ ID NO 765; 241pp; English.

XX CC The present invention describes a method of testing for bronchial asthma  
or chronic obstructive pulmonary disease. The method comprises  
determining the expression level of a marker gene in a biological sample  
from a subject, comparing the expression level determined with the  
expression level of the marker gene in a biological sample from a healthy  
subject, and judging whether the subject has bronchial asthma or chronic  
obstructive pulmonary disease. The marker gene comprises: (a) a group of  
genes (S1) whose expression levels increase when respiratory epithelial  
cells are stimulated with interleukin-13; or (b) a group of genes (S2)  
whose expression levels decrease when respiratory epithelial cells are

Aau05396 Human tit  
Adq17316 Human sof  
Adi21713 Rat LRP b  
Aae11937 Human CG1  
Abb05008 Mouse ree  
Abb57065 Mouse isc  
Ade60142 Rat Prote  
Adn17340 V. faba e  
Abb58665 Drosophila  
Abb11353 Human LDL  
Adn11590 Human CD9  
Adn11588 Human CD9  
Adn11587 Human CD9  
Adn11589 Human CD9  
Aar47861 Alpha 2-M  
Aar60517 Human alp  
Aam79091 Human pro  
Aau81019 Human alp  
Abp56839 Human LRP  
Abu89744 Protein d

26 2575.8 70.6 26926 4 AAU05396  
27 2575.8 70.6 26926 8 ADQ17316  
28 2575.1 70.6 4660 8 ADI21713  
29 2574.8 70.6 4636 4 AAE11937  
30 2571.1 70.5 3461 5 ABB05008  
31 2571.1 70.5 3461 5 ABB57065  
32 2571.1 70.5 3461 7 ADE60142  
33 2566.1 70.3 5825 8 ADN17340  
34 2564.8 70.3 4796 4 ABB58665  
35 2563.6 70.3 4563 4 ABB11353  
36 2562.6 70.2 4419 8 ADN11590  
37 2562.6 70.2 4419 8 ADN11588  
38 2562.6 70.2 4464 8 ADN11587  
39 2562.6 70.2 4464 8 ADN11589  
40 2562.6 70.2 4544 2 AAR47861  
41 2562.6 70.2 4544 2 AAR60517  
42 2562.6 70.2 4544 4 AAM79091  
43 2562.6 70.2 4544 5 AAU81019  
44 2562.6 70.2 4544 6 ABP56839  
45 2562.6 70.2 4544 6 ABU89744

CC stimulated with interleukin-13. Also described: (1) a reagent (I) for  
CC testing for bronchial asthma or chronic obstructive pulmonary disease;  
CC (2) a kit for screening for a candidate compound for a therapeutic agent  
CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)  
CC an animal model for bronchial asthma or chronic obstructive pulmonary  
CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a  
CC method for producing an animal model for bronchial asthma or chronic  
CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial  
CC asthma or chronic obstructive pulmonary disease, comprising the compound,  
CC a marker gene or an antisense nucleic acid corresponding to a portion of  
CC the marker gene, a ribozyme, a polynucleotide that suppresses the  
CC expression of the gene through an RNAi effect or an antibody recognising  
CC a protein encoded by a marker gene; and (7) a DNA chip for testing for  
CC bronchial asthma or a chronic obstructive pulmonary disease, on which a  
CC probe has been immobilised to assay a marker gene. (I) has respiratory  
CC and antiasthmatic activities, and can be used in gene therapy. The method  
CC is useful for testing for or screening for a therapeutic agent for  
CC bronchial asthma or chronic obstructive pulmonary disease. The present  
CC sequence is used in the exemplification of the present invention.

XX SQ Sequence 514 AA;

Query Match 94.2%; Score 3437; DB 8; Length 514;  
Best Local Similarity 90.4%; Pred. No. 8.9e-52;  
Matches 483; Conservative 6; Mismatches 5; Indels 40; Gaps 12;

QY 1 MSISSDEVNVLVYRYLOESGFSHSAFTFGIKSHISOSNINGALVPPAALISITIKGLQYV 60  
DB 1 MSISSDEVNVLVYRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISITIKGLQYV 60  
QY 61 EAEVSNEDGTLDPGRPIESLSLIDAVMPDVVOTQOAYRDKLAQOAAAAAASQ 120  
DB 61 EAEVSNEDGTLDPGRPIESLSLIDAVMPDVVOTQOAYRDKLAQOAAAAAASQ 120  
QY 121 QGSAKNGENTANGEANGAHTIANNHTDMEVDGVEIPPNKAVVLRGHESEVFCANVPV 180  
DB 121 QGSAKNGENTANGEANGAHTIANNHTDMEVDGVEIPPNKAVVLRGHESEVFCANVPV 180  
QY 181 SLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240  
DB 181 SLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240  
QY 241 ATGSDGDFARIWTKDGNLSTLQHGKPIFALKWNKKNFILSAGVDKTTIWDHTGEA 300  
DB 241 ATGSDGDFARIWTKDGNLSTLQHGKPIFALKWNKKNFILSAGVDKTTIWDHTGEA 300  
QY 301 KQOPPHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 360  
DB 301 KQOPPHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 360  
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDILQAHKEIYTIKWSPTGPTGNPNANLMLASAF 420  
DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDILQAHKEIYTIKWSPTGPTGNPNANLMLASAF 420  
QY 421 DSTVRLVDVDRGICHTLTKHQPVSVAFPDGRYLASGFDKCVHIMNTQ----VCLH 476  
DB 421 DSTVRLVDVDRGICHTLTKHQPVSVAFPDGRYLASGFDKCVHIMNTQ----VCLH 476  
QY 477 -Y--LNG--QVLN-----LGR-----SICLYTLPHHLVVIPLVALIELVL-K 514  
DB 479 SYRGTGTGIFECVNAAGDKVGSASGDSVC-----V-L-D---LRK 514

RESULT 2

AAB95225

ID AAB95225 standard; protein; 514 AA.

XX AC AAB95225;

XX AC AAB95225;

DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:17352.

XX

KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX Homo sapiens.  
XX EPI074617-A2.  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-00116126.  
XX PR 29-JUL-1999; 99JP-00248036.  
XX PR 27-AUG-1999; 99JP-00300253.  
XX PR 11-JAN-2000; 2000JP-00118776.  
XX PR 02-MAY-2000; 2000JP-00183767.  
XX PR 09-JUN-2000; 2000JP-00241899.  
XX (HELI-) HELIX RES INST.  
XX Ota T, Isogai T, Nishikawa T, Hayaashi K, Saito K, Yamamoto J;  
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.  
XX Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.  
XX The present invention describes primer sets for synthesising 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primers are useful for synthesising polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention

XX SQ Sequence 514 AA;

Query Match 93.9%; Score 3428; DB 4; Length 514;  
Best Local Similarity 90.3%; Pred. No. 1.3e-51;  
Matches 482; Conservative 6; Mismatches 6; Indels 40; Gaps 12;

QY 1 MSISSDEVNVLVYRYLOESGFSHSAFTFGIKSHISOSNINGALVPPAALISITIKGLQYV 60  
DB 1 MSISSDEVNVLVYRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISITIKGLQYV 60  
QY 61 EAEVSNEDGTLDPGRPIESLSLIDAVMPDVVOTQOAYRDKLAQOAAAAAASQ 120  
DB 61 EAEVSNEDGTLDPGRPIESLSLIDAVMPDVVOTQOAYRDKLAQOAAAAAASQ 120  
QY 121 QGSAKNGENTANGEANGAHTIANNHTDMEVDGVEIPPNKAVVLRGHESEVFCANVPV 180  
DB 121 QGSAKNGENTANGEANGAHTIANNHTDMEVDGVEIPPNKAVVLRGHESEVFCANVPV 180  
QY 181 SLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240  
DB 181 SLLASGSGDSTARIWNLSNSTSGTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240



Db 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQVPSNKDVTSLDWNSEGTLL 240  
QY 241 ATGSYDGFARIWTKDGNLSTLQGHKGPPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300  
Db 241 ATGSYDGFARIWTKDGNLSTLQGHKGPPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300  
QY 301 KQOFFPHSAPALDWDQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360  
Db 301 KQOFFPHSAPALDWDQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360  
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPTNNPNANMLASAF 420  
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPTNNPNANMLASAF 420  
QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQ----VCLH 476  
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQGTALV--H 478  
QY 477 -Y--LNG--QVLLN-----LGR-----SICLYTLPHHLVVIPLVALIELLVL-K 514  
Db 479 SYRGTGGIPEVCWNAAGDKVGASASDGSVC-----V-L-D-----LRK 514

RESULT 3  
ABO07190  
ID ABO07190 standard; protein; 514 AA.  
XX ABO07190;  
XX DT 13-AUG-2003 (first entry)  
XX DE Human p53 modifying protein, SEQ ID 150.  
XX KW Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;  
XX KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;  
XX KW lung cancer; ovarian cancer; angiogenesis; cell cycle;  
XX KW apoptotic disorder; cell proliferation disorder.  
XX OS Homo sapiens.  
XX XX  
XX WO200299122-A1.  
XX PN 12-DEC-2002.  
XX PD 03-JUN-2002; 2002WO-US017382.  
XX PF 05-JUN-2001; 2001US-0296076P.  
XX PR 10-OCT-2001; 2001US-0328605P.  
XX PR 15-FEB-2002; 2002US-0357253P.  
XX XX  
XX PA (EXEL-) EXELIXIS INC.  
XX XX  
XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;  
XX PI  
XX WP1: 2003-156859/15.  
XX DR N-PSDB; ACD13365.  
XX DR  
XX XX  
XX Identifying modulators of the p53 pathway for use in treating apoptotic  
XX PT or cell proliferation disorders, comprises screening for agents that  
XX PT modulate activity of a human ortholog of genes that modify the p53  
XX PT pathway in Drosophila.  
XX XX  
XX Example 2; Page 458-459; 678pp; English.  
XX XX

The invention relates to identifying (M1) a candidate p53 pathway  
modulating agent, by contacting an assay system comprising a purified HM  
polypeptide (human orthologue of genes that modify the p53 pathway in  
Drosophila) or nucleic acid with a test agent under conditions, where but  
for the presence of the test agent, the system provides a reference  
activity, and detecting a test agent-biased activity of the assay system.  
Also included are modulating (M2) a p53 pathway of a cell (comprising  
contacting a cell defective in p53 function with a candidate modulator  
that specifically binds to a HM polypeptide comprising an HM amino acid

CC sequence, where p53 function is restored), modulating (M3) a p53 pathway  
in a mammalian cell (comprising contacting the cell with an agent that  
specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)  
a disease in a patient (comprising: (a) obtaining a biological sample  
from the patient; (b) contacting the sample with a probe for HM  
expression; (c) comparing the results with a control; and (d) determining  
whether the comparison indicates a likelihood disease). (M1) is useful  
for identifying modulators of the p53 pathway. A probe for HM expression  
is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,  
in a patient, where the cancer has greater than 25 % expression level.  
Modulators identified by (M1) are useful in a variety of diagnostic and  
therapeutic applications, where disease or disorder prognosis is related  
to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
proliferation disorders (e.g. cancer). Another two new methods (M2 and  
M3) are useful for modulating the p53 pathway of a cell, thus restoring  
the p53 function of the cell, so that the cell undergoes normal  
proliferation or progression through the cell cycle. (M2) and (M3) are  
also useful for treating defects in the p53 pathway such as angiogenic,  
apoptotic or cell proliferation disorders. The present sequence  
represents a human p53 pathway modifying protein

XX  
SQ Sequence 514 AA;

Query Match 93.9%; Score 3427; DB 6; Length 514;  
Best Local Similarity 90.3%; Pred. No. 1.4e-51;  
Matches 482; Conservative 6; Mismatches 6; Indels 40; Gaps 12;

QY 1 MTSDEVNFLVRYLQESGFHSFTFGIKSHISOSNINGALVPPAALISIIQKGLQYV 60  
Db 1 MTSDEVNFLVRYLQESGFHSFTFGIKSHISOSNINGALVPPAALISIIQKGLQYV 60  
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQQAAAAAASQ 120  
Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQQAAAAAASQ 120  
QY 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFI 180  
Db 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFI 180  
QY 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQVPSNKDVTSLDWNSEGTLL 240  
Db 181 SLLASGSGDSTARINWLSNSTSGTQLVLRHCIREGGQVPSNKDVTSLDWNSEGTLL 240  
QY 241 ATGSYDGFARIWTKDGNLSTLQGHKGPPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300  
Db 241 ATGSYDGFARIWTKDGNLSTLQGHKGPPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300  
QY 301 KQOFFPHSAPALDWDQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360  
Db 301 KQOFFPHSAPALDWDQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360  
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPTNNPNANMLASAF 420  
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQOHNKEIYTIKWSPTGPTNNPNANMLASAF 420  
QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQ----VCLH 476  
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQGTALV--H 478  
QY 477 -Y--LNG--QVLLN-----LGR-----SICLYTLPHHLVVIPLVALIELLVL-K 514  
Db 479 SYRGTGGIPEVCWNAAGDKVGASASDGSVC-----V-L-D-----LRK 514

RESULT 4  
ADJ76263  
ID ADJ76263 standard; protein; 514 AA.  
XX  
XX AC ADJ76263;  
XX DT 20-MAY-2004 (first entry)  
XX  
XX DE Marker gene related amino acid sequence SEQ ID NO:1515.

XX bronchial asthma; chronic obstructive pulmonary disease;  
 KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;  
 KW gene therapy; marker.  
 XX Mus musculus.  
 XX EPI394274-A2.  
 XX 03-MAR-2004.  
 XX 04-AUG-2003; 2003EP-00254857.  
 XX 06-AUG-2002; 2002JP-00229312.  
 XX 20-MAR-2003; 2003JP-00077212.  
 XX (GENO-) GENOX RES INC.  
 XX Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;  
 XX WPI; 2004-193155/19.  
 XX Testing for bronchial asthma or chronic obstructive pulmonary disease by  
 PT comparing the expression level of a marker gene in a biological sample  
 PT from a subject with the expression level of the gene in a sample from a  
 PT healthy subject.  
 XX Claim 16; SEQ ID NO 1515; 241pp; English.

XX The present invention describes a method of testing for bronchial asthma  
 CC or chronic obstructive pulmonary disease. The method comprises  
 CC determining the expression level of a marker gene in a biological sample  
 CC from a subject, comparing the expression level determined with the  
 CC expression level of the marker gene in a biological sample from a healthy  
 CC subject, and judging whether the subject has bronchial asthma or chronic  
 CC obstructive pulmonary disease. The marker gene comprises: (a) a group of  
 CC genes (S1) whose expression levels increase when respiratory epithelial  
 CC cells are stimulated with interleukin-13; or (b) a group of genes (S2)  
 CC whose expression levels decrease when respiratory epithelial cells are  
 CC stimulated with interleukin-13. Also described: (1) a reagent (I) for  
 CC testing for bronchial asthma or chronic obstructive pulmonary disease;  
 CC (2) a kit for screening for a candidate compound for a therapeutic agent  
 CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)  
 CC an animal model for bronchial asthma or chronic obstructive pulmonary  
 CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a  
 CC method for producing an animal model for bronchial asthma or chronic  
 CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial  
 CC asthma or chronic obstructive pulmonary disease, comprising the compound,  
 CC a marker gene or an antisense nucleic acid corresponding to a portion of  
 CC the marker gene, a ribozyme, a polynucleotide that suppresses the  
 CC expression of the gene through an RNAi effect or an antibody recognising  
 CC a protein encoded by a marker gene; and (7) a DNA chip for testing for  
 CC bronchial asthma or a chronic obstructive pulmonary disease, on which a  
 CC probe has been immobilised to assay a marker gene. (I) has respiratory  
 CC and antiasthmatic activities, and can be used in gene therapy. The method  
 CC is useful for testing for or screening for a therapeutic agent for  
 CC bronchial asthma or chronic obstructive pulmonary disease. The present  
 CC sequence is used in the exemplification of the present invention.

XX Sequence 514 AA;

Query Match 93.2%; Score 3402; DB 8; Length 514;  
 Best Local Similarity 89.3%; Pred. No. 4.2e-51;  
 Matches 477; Conservative 8; Mismatches 9; Indels 40; Gaps 12;

QY 1 MSISSEVNLVRYLQESGFSHSAFTFGIKSHISQNSINGALVPPAALISITKGLQYV 60  
 DB 1 MSISSEVNLVRYLQESGFSHSAFTFGIKSHISQNSINGALVPPAALISITKGLQYV 60  
 QY 61 EAEVSNEDCTLFDGRIEISLIDAVMPDVVQTRQAVEDKLAQQAQAAAAAASQ 120  
 DB 61 EAEVSNEDCTLFDGRIEISLIDAVMPDVVQTRQAVEDKLAQQAQAAAAAATNQ 120

QY 121 QGSAKXGENTANGEAGHTIANNHTDMMVEGDVEIPNKA VVLRGHESEVFCANPV 180  
 DB 121 QGSAKXGENTANGEAGHTIANNHTDMMVEGDVEIPNKA VVLRGHESEVFCANPV 180  
 QY 181 SLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240  
 DB 181 SLLVSGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240  
 QY 241 ATGSDGFARIWTKDGNLSTLQHGKPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300  
 DB 241 ATGSDGFARIWTKDGNLSTLQHGKPIFALKWKKGNFILSAGVDKTTIIWDAHTGEA 300  
 QY 301 KQOFFPHSAPALVDVQSNNTFASCSTDMCIHVCKLQDRPIKTFQGHTEVNAIKWDPT 360  
 DB 301 KQOFFPHSAPALVDVQSNNTFASCSTDMCIHVCKLQDRPIKTFQGHTEVNAIKWDPT 360  
 QY 361 GNLLASCSDDMTLKIWMSKQDNCVHDLQOHNKEIYTIKWSPTGPTGNNPNANMLASASF 420  
 DB 361 GNLLASCSDDMTLKIWMSKQDNCVHDLQOHNKEIYTIKWSPTGPTGNNPNANMLASASF 420  
 QY 421 DSTVRLWDVDRGICIHITLTKHOEPVYSAFSPDGRYLASGSPDKCVHIWNTQ----VCLH 476  
 DB 421 DSTVRLWDVDRGICIHITLTKHOEPVYSAFSPDGRYLASGSPDKCVHIWNTQ--H 478  
 QY 477 -Y--LNG--QVLIN-----LGR-----SICLYTLPHHLVVIPLVALIELVL-K 514  
 DB 479 SYRGTGGEIPECVWNAAGDKVASASDGSVC-----V-L-D-----LRK 514

RESULT 5  
 ADD14051

ID ADD14051 standard; protein; 577 AA.

AC ADD14051;

DT 01-JAN-2004 (first entry)

XX Human src biomarker polypeptide SEQ ID NO:240.

XX predictor set; protein tyrosine kinase activity modulator;

KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;  
 KW gene therapy; drug sensitivity; genetic profile; cancer; human.

OS Homo sapiens.

PN WO2003062395-A2.

XX 31-JUL-2003.

XX 17-JAN-2003; 2003WO-US001981.

XX 18-JAN-2002; 2002US-0350061P.

XX (BRIM ) BRISTOL-MYERS SQUIBB CO.

XX Huang F, Fairchild CR, Lee FY, Shaw P;

XX WPI; 2003-636735/60.

XX N-PSDB; ADD14646.

XX New polynucleotides and polypeptides for predicting the activity of  
 PT compounds that interact with protein tyrosine kinases and/or protein  
 PT tyrosine kinase pathways.

XX Claim 10; SEQ ID NO 240; 139pp; English.

XX The present invention describes a predictor set comprising a plurality of  
 CC polynucleotides or polypeptides whose expression pattern is predictive of  
 CC the response of cells to treatment with a compound that modulates protein  
 CC tyrosine kinase activity or members of the protein tyrosine kinase  
 CC pathway. Also described: (1) predicting whether a compound is capable of  
 CC modulating the activity of cells, comprising obtaining a sample of cells,  
 CC determining whether the cells express a plurality of markers, and





08-NOV-2000; 2000US-0246609P.  
08-NOV-2000; 2000US-0246610P.  
08-NOV-2000; 2000US-0246611P.  
08-NOV-2000; 2000US-0246613P.  
17-NOV-2000; 2000US-0249207P.  
17-NOV-2000; 2000US-0249208P.  
17-NOV-2000; 2000US-0249209P.  
17-NOV-2000; 2000US-0249210P.  
17-NOV-2000; 2000US-0249211P.  
17-NOV-2000; 2000US-0249212P.  
17-NOV-2000; 2000US-0249213P.  
17-NOV-2000; 2000US-0249214P.  
17-NOV-2000; 2000US-0249215P.  
17-NOV-2000; 2000US-0249216P.  
17-NOV-2000; 2000US-0249217P.  
17-NOV-2000; 2000US-0249218P.  
17-NOV-2000; 2000US-0249244P.  
17-NOV-2000; 2000US-0249245P.  
17-NOV-2000; 2000US-0249264P.  
17-NOV-2000; 2000US-0249265P.  
17-NOV-2000; 2000US-0249297P.  
17-NOV-2000; 2000US-0249299P.  
17-NOV-2000; 2000US-0249300P.  
01-DEC-2000; 2000US-0250160P.  
01-DEC-2000; 2000US-0250391P.  
05-DEC-2000; 2000US-0251030P.  
05-DEC-2000; 2000US-0251988P.  
05-DEC-2000; 2000US-0256719P.  
06-DEC-2000; 2000US-0251479P.  
08-DEC-2000; 2000US-0251856P.  
08-DEC-2000; 2000US-0251868P.  
08-DEC-2000; 2000US-0251869P.  
08-DEC-2000; 2000US-0251989P.  
08-DEC-2000; 2000US-0251990P.  
11-DEC-2000; 2000US-0254097P.  
05-JAN-2001; 2001US-0259678P.  
XX  
(HUMA-) HUMAN GENOME SCI INC.  
XX  
Rosen CA, Barash SC, Ruben SM;  
XX  
WPI; 2001-476159/51.  
DR N-PSDB; ADM19381.  
XX  
Isolated nucleic acid molecule encoding a channel/transporter protein is  
PT used in preventing, treating or ameliorating a medical condition.  
XX  
Claim 11; SEQ ID NO 667; 809pp; English.  
XX  
The invention relates to an isolated nucleic acid molecule encoding a  
CC channel/transporter protein or sequences at least 95% identical to a  
CC these. The nucleic acids and proteins encoded by them are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
CC in diagnosing a pathological condition or susceptibility to a  
CC pathological condition. The antibodies to the proteins can also be used  
CC in alleviating symptoms associated with the disorders and in diagnostic  
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
CC (ELISA). Disorders which are diagnosed or treated include autoimmune  
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,  
CC nervous system disorders e.g. Alzheimer's disease, infections caused by  
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.  
CC The polypeptides can also be used to aid wound healing and epithelial  
CC cell proliferation, to prevent skin aging due to sunburn, to maintain  
CC organs before transplantation, for supporting cell culture of primary  
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can  
CC also be used as a food additive or preservative to increase or decrease  
CC storage capabilities. This sequence corresponds to a protein of the  
CC invention.  
XX  
Sequence 459 AA;  
SQ

Query Match 84.1%; Score 3069.7; DB 4; Length 459;  
Best Local Similarity 83.6%; Pred. No. 5.4e-45;  
Matches 428; Conservative 5; Mismatches 5; Indels 73; Gaps 15;  
QY 23 HSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYVEAEVSEINEDGTLFDGRPIESLS 82  
DB 1 H-A-----SD-----VEAEVSEINEDGTLFDGRPIESLS 27  
QY 83 LIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQGSAGKNGTANGENGATIA 142  
DB 28 LIDAVMPDVVQTRQAYRDKLAQOQAAAAAASQGSAGKNGTANGENGATIA 87  
QY 143 NNHTDMMEYDGDVEIPPNKAVLVRGHESEVFICANNPVSDLLASGSDSTARIWNLSSENS 202  
DB 88 NNHTDMMEYDGDVEIPPNKAVLVRGHESEVFICANNPVSDLLASGSDSTARIWNLSSENS 147  
QY 203 TSGSTQLVLRHCIREGGQDVPSNKQVTSLOWNSEGTLATGSDYDGFARITWKDGNLASTL 262  
DB 148 TSGSTQLVLRHCIREGGQDVPSNKQVTSLOWNSEGTLATGSDYDGFARITWKDGNLASTL 207  
QY 263 GQHKGPFAKWNKKNFILSAGVDKTTIWDATGCAKQPPHSAAPALDQVQSNNTF 322  
DB 208 GQHKGPFAKWNKKNFILSAGVDKTTIWDATGCAKQPPHSAAPALDQVQSNNTF 267  
QY 323 ASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGNTNLLASCDDMTLKIWSMKQDN 382  
DB 268 ASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGNTNLLASCDDMTLKIWSMKQDN 327  
QY 383 CVHDLQHNKEIYTIKWSPGTGPTNNPNANMLASFPDSTVRLVDVDRGICHTITKHQ 442  
DB 328 CVHDLQHNKEIYTIKWSPGTGPTNNPNANMLASFPDSTVRLVDVDRGICHTITKHQ 387  
QY 443 EPYISVAFSPDGRYLASGSDKCVHIWNTQ-----VCLH-Y--LNG--QVLLN-----LGR 488  
DB 388 EPYISVAFSPDGRYLASGSDKCVHIWNTQ-----VCLH-Y--LNG--QVLLN-----LGR 445  
QY 489 -----SICLYTLPHHLVWIPLVALIELVL-K 514  
DB 446 SASDGSVC-----V-L-D---LRK 459  
RESULT 8  
ABP41760  
ID ABP41760 standard; protein; 542 AA.  
XX  
AC ABP41760;  
XX  
DT 22-AUG-2002 (first entry)  
XX  
DE Human ovarian antigen HNOXK38, SEQ ID NO:2892.  
XX  
KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
KW inflammatory condition; immune disorder; blood disorder;  
KW cardiovascular disorder; respiratory disorder; neurological disorder;  
KW gastrointestinal disorder; urinary system disorder; drug screening;  
KW gene therapy; chromosome mapping; forensic analysis;  
KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
KW antiinflammatory; gynaecological; reproductive.  
OS Homo sapiens.  
XX  
XX WO200200677-A1.  
PN  
XX  
PD 03-JAN-2002.  
XX  
XX 07-JUN-2001; 2001WO-US018569.  
PF  
XX  
XX 07-JUN-2000; 2000US-0209467P.  
PR  
XX



Db 1 MSFSDSEVNLVRYLQESGFLSHAVVGFIESHSQNSINGALVPAALITLQKGLYT 60  
QY 61 EAEVSNEDGTLFD-GRPIESLSLIDAVMPDV-----VOTROQ-----AYRD---KLA-- 104  
Db 61 EVESVGEDG---EVARPIEGLSLIDAVMPDKPIVKT-EPGKPGAV-DSSAP-AGG 114  
QY 105 -Q-----Q-----Q-A-----AA-A-----A 112  
Db 115 NONNAKPEIKIEPGTGVAGSAGNKIAGSTTGTSTPTDQSAASEVDSSGNAANAGGTYA 174  
QY 113 ----A-----A-----AA-ASQ-----Q-----GSAK--N-G-EN---T-- 130  
Db 175 GNNAGGNOASTGSGNSTSTPAGDGLAAPGASOKSQNSNEAGSSGSGNAGNANATSTD 234  
QY 131 -A-----NG-----EE-----N-----G-AH-----T- 141  
Db 235 AASSTNGNSTSSSVEQTSGLTTPAGGTVSTNPDAAAGSAGSTATSGKAGGAVTIR 294  
QY 142 --A--NN-----H-----T-DMVEV--DGD--VEIPPNKAVVLRG 167  
Db 295 VQAQGNVQSGSSNAQSSAPSGTSSSTSGGAGTTPAAL-VPMDIDENIEIPESKARVLG 353  
QY 168 HESEVFIQAWNPVS-DLLASGSGDSTARINLSE-NSTGSGTOLVLRHCIREGGQDVPSN 225  
Db 354 HESEVFIQAWNP-SRDLASGSGDSTARINWMSDAN-T-NSNQLVLRHCITQKGAEPVSN 410  
QY 226 KQVTSLDWNSGTLTATGSDGPARVW-TWQGNLASTLQGHKPIALKNKK-GNFILS 283  
Db 411 KQVTSLDWNCDSGLLATGSDYGVARVWKT-DGRLASTLQGHKPIALKNW-KCGNYILS 468  
QY 284 AGVDKTTIWDATGE-AKQPFPHSAPALDVQSNNTFASGSTD-MCHVCKLG--OD 339  
Db 469 AGVDKTTIWDASTGQCT-QQFAFHPALDVQSNNTFASGSTDOR-IHVCRLGWE- 525  
QY 340 RPIKTFQHTNEVNAIKWPTGNLLASCDSDMTLKIWSMKQD-NCVHDLQOHNKIYTIK 398  
Db 526 -PIKTFKHTNEVNAIKWCPQGLLASCDSDMTLKIWSMNRDCC-HDLQAHSEIYTIK 583  
QY 399 WSPGTGTPNPNANLASAFSTVRLWDVDRGICHTLTKHOEPVSVASPDGRYLA 458  
Db 584 WSPGTGTPNPNANLASAFSTVRLWDVDRGICHTLTKHOEPVSVASPDGRYLA 643  
QY 459 SGSDFKCVHIWNTQ-----VCLH-YLNGQVLLNLGRSI---CLY-----TLPHLV----- 500  
Db 644 SGSDFKCVHIWNTQGLV--HSY-KG-----TG-GIFEVC-WNSKGT-K---VGASAD 689  
QY 501 --VPLVALIELVL-K 514  
Db 690 GSV--FV-L-D---LRK 699  
RESULT 10  
ABP51424  
ID ABP51424 standard; protein; 395 AA.  
XX  
AC ABP51424;  
XX  
DT 03-SEP-2002 (first entry)  
DE  
DE Human MDDT SEQ ID NO 446.  
XX  
KW Human; MDDT; disease detection and treatment molecule polynucleotide;  
KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;  
KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;  
KW antiallergic; antianaemic; antiasthmatic; antiarteriosclerotic; antitumor;  
KW neuroprotective; antirheumatic; antiarthritic.  
OS Homo sapiens.  
XX  
XX W0200240715-A2.

XX  
PD 23-MAY-2002.  
XX  
PF 06-SEP-2001; 2001WO-US027628.  
XX  
PR 05-SEP-2000; 2000US-0229747P.  
PR 05-SEP-2000; 2000US-0229748P.  
PR 05-SEP-2000; 2000US-0229749P.  
PR 05-SEP-2000; 2000US-0229750P.  
PR 05-SEP-2000; 2000US-0229751P.  
PR 05-SEP-2000; 2000US-0230583P.  
PR 06-SEP-2000; 2000US-0230505P.  
PR 06-SEP-2000; 2000US-0230514P.  
PR 06-SEP-2000; 2000US-0230515P.  
PR 06-SEP-2000; 2000US-0230517P.  
PR 06-SEP-2000; 2000US-0230518P.  
PR 06-SEP-2000; 2000US-0230519P.  
PR 06-SEP-2000; 2000US-0230595P.  
PR 06-SEP-2000; 2000US-0230597P.  
PR 06-SEP-2000; 2000US-0230598P.  
PR 06-SEP-2000; 2000US-0230599P.  
PR 06-SEP-2000; 2000US-0230610P.  
PR 06-SEP-2000; 2000US-0230865P.  
PR 06-SEP-2000; 2000US-0230988P.  
PR 07-SEP-2000; 2000US-0230951P.  
PR 07-SEP-2000; 2000US-0231163P.  
PR 07-SEP-2000; 2000US-0231167P.  
XX  
XX (INCY-) INCYTE GENOMICS INC.  
XX  
PI Jackson S, Lincoln SB, Altus CM, Dufour GE, Chalup MS;  
PI Hillman JL, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE;  
PI Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B,  
PI Roseberry AM, Gerstlin EH, Peralta CH, David MH, Panzer SR, Flores V;  
PI Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
XX  
XX WPI; 2002-527544/56.  
XX N-PSDB; ABQ72641.  
XX  
XX Novel human disease detection and treatment polypeptide, useful in  
XX diagnosis, prevention or treatment of cell proliferative disorders e.g.  
XX arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.  
XX AIDS.  
XX  
XX Claim 14; Page 578; 618pp; English.  
XX  
XX The invention relates to an isolated human disease detection and  
XX treatment (MDDT) polypeptide (I) selected from a polypeptide having a  
XX sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
XX specification, a naturally occurring polypeptide comprising a sequence  
XX having at least 90% identity to (I) or a biologically active or  
XX immunogenic fragment of (I). (I) is useful for screening a compound for  
XX effectiveness as an agonist or antagonist, for screening a compound that  
XX specifically binds (I) or modulates the activity of (I), and for  
XX preparing a polyclonal or monoclonal antibody by hybridoma technology.  
XX Nucleic acids (II) (ABQ7249-ABQ72700) encoding (I) are useful for  
XX screening a compound for effectiveness in altering expression of a target  
XX polynucleotide comprising. Oligonucleotides and antibodies are useful for  
XX detecting MDDT in a sample or for assessing toxicity of a test compound,  
XX in a diagnostic test for a condition or a disease associated with the  
XX expression of MDDT in a biological sample, for detecting (I) in a sample,  
XX and for purifying (I) from a sample. A composition comprising (I), an  
XX agonist or antagonist is useful for treating a disease or condition  
XX associated with decreased or increased expression of functional MDDT. (I)  
XX or (II) are useful for diagnosing, treating or preventing disorders  
XX associated with aberrant expression of MDDT, where the disorders are  
XX selected from a cell proliferative disorder such as arteriosclerosis,  
XX cirrhosis, hepatitis, psoriasis, and cancer and an  
XX autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
XX allergy, anaemia, asthma, arteriosclerosis, gout, multiple sclerosis or  
XX rheumatoid arthritis. (II) are useful for creating knockin humanised  
XX animals or transgenic animals to model human diseases, in somatic or

```
CC germline gene therapy, to generate a transcript image of a tissue or cell
CC type, for detecting differences in the chromosomal location due to
CC translocation or inversion among normal, carrier or affected individuals
CC and as hybridisation probes for mapping naturally occurring genomic
CC sequences
XX
SQ Sequence 395 AA;
Query Match 75.2%; Score 2744; DB 5; Length 395;
Best Local Similarity 99.5%; Pred. No. 4.4e-39;
Matches 389; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSISSDEVNVLVRYIQESGFSHSAFTFGIKSHISOSNINGALVPPAALISIIQKGLQV 60
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 5 MSISSDEVNVLVRYIQESGFSHSAFTFGIESHISOSNINGALVPPAALISIIQKGLQV 64
QY 61 EAEVSNEDCTLDFGRPIESLSLIDAVMPDVQTRQAYRDKLAQQQAAAAAASQ 120
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 65 EAEVSNEDCTLDFGRPIESLSLIDAVMPDVQTRQAYRDKLAQQQAAAAAASQ 124
QY 121 QGSAKNGENTANGENGATIANHTDMMEVDGVEIPPNKAVLGRHSEVEFCANVP 180
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 125 QGSAKNGENTANGENGATIANHTDMMEVDGVEIPPNKAVLGRHSEVEFCANVP 184
QY 181 SLLASGSGDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 185 SLLASGSGDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 244
QY 241 ATGSDGCFARIWTKDGNLSTLQGHKGPFPALKWNKKNFILSAGVDKTTIWDHTGEA 300
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 245 ATGSDGCFARIWTKDGNLSTLQGHKGPFPALKWNKKNFILSAGVDKTTIWDHTGEA 304
QY 301 KQPPFHSAPALVDVWQSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 360
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 305 KQPPFHSAPALVDVWQSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 364
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHN 391
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 365 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHN 395
RESULT 11
ABG21351
ID ABG21351 standard; protein; 584 AA.
XX
AC ABG21351;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #21342.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
FN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US008631.
XX
PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS85538.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
```

```
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 51710; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities in
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG3037 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 584 AA;
Query Match 72.4%; Score 2643.1; DB 4; Length 584;
Best Local Similarity 58.1%; Pred. No. 2.4e-36;
Matches 364; Conservative 45; Mismatches 39; Indels 179; Gaps 63;
QY 2 SI-----S-----S-----D-----EV---N---FLV-YEY-LQE-SGRSHSAFTFGIKSH 33
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 22 AIGAVGFRSQAASTLAGQPDGNGVCVHAGNSRIF-VLPVGL-EVKGFSHSAFTFGIESH 79
QY 34 ISQSNINGALVPPAALISIIQKGLQVVEAEVSNEDGTLFDGRPIESLSLIDAVMPDVQ 93
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 80 ISQSNINGTLVPPSALISILQKGLQVVEAFISINKDGTVDSPRIESLSLIVAVIPDVQ 139
QY 94 TRQAYRDKLAQQQA-AAA-A-A-AA-----AA-SQGSAXKNGENTANGENGATTA 142
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 140 MRQQAFFGEKLTQQAASAAATEASAKAATMTPAASIQSNPPKNEATVNGEENGAHEI- 198
QY 143 NNHTDM-MEVDGVEIPPNKAVLGRHSEVEFCANVPVSDLLASGSGDSTARINWLSN 201
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 199 NNHS-KPMEIDGVEIPPNKATVLRHSEVEFCANVPVSDLLASGSGDSTARINWLNEN 257
QY 202 STSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLLATGSDGFARITKDGKLNAST 261
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 258 SNGSGTQLVLRHCIREGGHVDPSNKDVTSLDWNSDGTLAMGSDYDGFARITWE--N-A-- 312
QY 262 LGQHKGPALP--KW-NKKGN--FILS-AGVDKTTIWDHT--GEAKQOFFFHSAPAL 312
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 313 -----P--ALDVDWQN---NMTFA-SCS-TDM--CI---HVCRLGCD-----H--P-- 344
QY 313 DVD-WQSNNTFASCSTDMCIH-VC-----KL-----G-QDRPIKTF-----QG- 347
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 345 -VKTFO-GHTF--C-T--CIESICFWGGLRKLTTWTEGKLRP-KTFCSDGGALLPAGR 396
QY 348 --H--T-----NEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQ 389
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 397 RPHLLTGPDIFFKRLSALFQNEVNAIKWDPGSMLLASCSDDMTLKIWSMKQDNCVHDLQ 456
QY 390 HNKEIYTIKWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTKHOBPVYSVA 449
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 457 HSKEIYTIKWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTKHOBPVYSVA 516
QY 450 FSPDGRYLASGSPDKCVHIWNTQ---VCLH--Y--LNG--QVLLN-----LGR-----SI 490
DB |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 517 FSPDGKYLASGSPDKCVHIWNTQSGSLV--HSYQGTGGEFECVNCNARGDKVGSASGVS 574
```



QY	491 C-LYTLPHLWVPLVALIELVL--K 514	Db	1492 QPLQVHDIOKESWINTYSGRLNRQSLQLQVDSFLRDPSPSSVAVVAVCNIPYSMTYVERHV 1551
Db	575 CILW-----Q---LSGK 583	QY	15 -----YL-----OE-S-----G-----FS----- 22
RESULT 12		Db	1552 IEWLDSGLTAGDAEDWLSIROTSECSLSALDLOQIAHOTGMQVEIWSRQFSORGG 1611
ADG70546		QY	23 -----HS-----A-F-T-----F-----G-----I-----K 31
ADG70546 standard; protein; 7064 AA.		Db	1612 LDAIFHRHSGDRTSRALENFPTDYQGRPFQSLSKWPLQRQGEKQITLGDVSTVCK 1671
ADG70546;		QY	32 -----S-----H-IS-----QS-NI-----NG-----A 42
11-MAR-2004 (first entry)		Db	1672 EDLHDITWNEVVPDALEACVHDLISDTVRAQSPAIACAWDGEWSYIELDLSSRLAHA 1731
Aspergillus niger nsb2 gene-encoded peptide synthetase.		QY	43 L-----VP-----P-----AA-----L-I-SII-Q----- 54
Ferrichrome biosynthesis gene cluster; nsb2; peptide synthetase;		Db	1732 LAPFGVANTVVICPEKSKWTPVATLAVMKAGAASVTLDASQPLERLRSIIQTDPRVIL 1791
transgenic; ferrichrome production; siderophore; iron chelating agent;		QY	55 -----K-GL-QY-----V-E-----AB-----V- 64
black koji; malt; food product; beverage; pharmaceutical;		Db	1792 SSASKQGLGAQLTKAPNLVVDQHSISTMHTAEPPLTVDPSSKLYIVFTSGTGVKGI 1851
functional food; nutraceutical; anaemia; sideropenia; ischaemia;		QY	65 -----S-I-----N-----ED----- 69
antianemic; vasotropic; enzyme.		Db	1852 ITHSNFSSAIRHQKAGFKSTRIYDFASYAFDVSNFIHALTVGACLICPSDEDRD 1911
Aspergillus niger.		QY	70 ---GTL--F---D-----G-----R----- 76
JP2003230381-A.		Db	1912 DLAGSLERFATHVDMTPSAASLLPEKSPKRLTETVVLGGELKSVESAQRWSSLSLVKNPY 1971
19-AUG-2003.		QY	77 -----P-I-----E-S-----L-----SLI----- 84
06-FEB-2002; 2002JP-00030145.		Db	1972 GPSECTPTATVPTDEYKSGISRGGLNTWNTVTDLSLVPVGVGGLLEGLPLGA 2031
06-FEB-2002; 2002JP-00030145.		QY	85 -----D-----A-V-----M-----PD----- 90
(GEKK-) GEKKEIRAN KK.		Db	2032 GYLGDTKTAASFEVDPQLLQICQOQARHTRMYKTGDLVHVPDGSLSFVGRKDAQVK 2091
WPI; 2003-820206/77.		QY	91 -----V-----VQ-----T-----R-----Q-----Q----- 97
N-PSDB; ADG70547.		Db	2092 IHQORVELTEIESHIRTSTKIQAVLFTKSGLCANRVAVFCIOGTGTQTAADIRLI 2151
Novel ferrichrome biosynthesis cluster comprising ornithine monooxygenase		QY	98 -----AY-----R----- 100
protein and peptide-synthetase protein, useful for producing ferrichrome		Db	2152 DPYSTLVITAYTESAKSSLDLTPAYMIPSIWIPLOHVPLSTSGKLDYKALKSWLDSMDA 2211
utilized in food or beverage products and in pharmaceuticals.		QY	101 -----D-KL--A---QQ--QA-A---A---A---A---A---A---A--- 112
Claim 2; SEQ ID NO 3; 60pp; Japanese.		Db	2212 KTFANILTASDGDVLRKAETELEQVIVEACAKILNITASKVNLDRSFANGGDSISAMR 2271
The invention relates to the Aspergillus niger ferrichrome biosynthesis		QY	113 --A--A---A---AA-----AS----- 119
gene cluster which comprises the nsb1 gene (ADG70545) encoding an		Db	2272 LVAHCRADVNVFSAKLKSKTLAALASSKIKSASNVILGFVEEKSDFSALSPIOQWFFE 2331
ornithine monooxygenase (L-ornithine N5-oxygenase; ADG70544) and the nsb2		QY	120 Q-----QG-----SA-K-----NG-E---NT--- 130
gene (ADG70547) encoding a peptide synthetase (ADG70546). The invention		Db	2332 QGLYKRSNDNFQGFYLVKKEPLLTDKDIDSAISKVQVHSMRLARFHRNGDEWTKTLK 2391
also relates to recombinant vectors, host cells and transgenic organisms		QY	131 --ANG-----E-----E-----N--G-----AH-- 139
comprising at least one of the ferrichrome biosynthesis gene cluster		Db	2392 PDTNGLYHPGVHHTCLPADIERLALSRHQMDIEKGPVFSADICHNAFGEQVILIAHL 2451
coding regions; and method for protein engineering of a peptide		QY	140 ----- 144
synthetase involving domain rearrangement. Transgenic organisms		Db	2452 VVDLSVRVILEDBSLGGSNLQPSLPFQVWMDMQIERAKESSLDIPENVLSTGINNN 2511
comprising all or part of the ferrichrome biosynthesis gene cluster		QY	145 -----H--T--D-----M-----ME-VD----- 152
(e.g., Escherichia coli or Aspergillus sp.) are able to produce large		Db	2512 LDFWQATATKMTVEDHLNFCYKIDSSKSSSKTKITRSTLEPVDLLLAHVHAFKTF 2571
amounts of ferrichrome using black koji (malt) as a substrate.		QY	153 -----G-----DVE-----I-P----- 158
Ferrichrome is a safe iron-chelating agent (siderophore) which is widely		Db	2572 PQDGLTIFIEGHGEPWSSDIDLSTRTVGNFTTISPIHVSKSDVHKSVASLVVRVKDARR 2631
used in pharmaceuticals or in food or beverage products. It can be used			
for the amelioration of anaemia (particularly in a functional food) and			
for treating sideropenia and ischaemia. The present sequence represents			
the Aspergillus niger ferrichrome biosynthesis gene cluster nsb2 gene-			
encoded peptide synthetase.			
Sequence 7064 AA;			
Query Match	71.4%; Score 2605.2; DB 7; Length 7064;		
Best Local Similarity	9.1%; Pred. No. 2.2e-30;		
Matches	453; Conservative 48; Mismatches 6; Indels 4448; Gaps 368;		
QY	1 M-----S-----I-----S-----S----- 5		
Db	1432 MEQVELELLVDPAFTSLVDPRPGLVEHVEILLPKRLKSTNELSAYRYAAVHLKDSNQLA 1491		
QY	6 -----D---E--VN-----FL-----VY-R-- 14		

QY 159 --P-----N-KA--V--V-- 164  
Db 2632 LLPANGWAYPASRYFNESGSAFKSHDSIMEITFNHYGFOQLENEKAMPENVTLSGVCE 2691  
QY 165 ---L---RG---H---E-S-E 171  
Db 2692 QGPALPASSLIAVEVSIDRQVTFDVSANRYINHQCISNWIKAISQSLETISNELVSTE 2751  
QY 172 V-----F-----I-C----- 175  
Db 2752 ISHRTLCDVEFLSLGVTEDRLQESVPIETEKLNNSTVECIYRCLPTVDGILISQFKDPE 2811  
QY 176 ---AW---N---P-----V--- 180  
Db 2812 SYKTVOHFETSHIDDQIDLEHLSLAWQKVANQPALRTVFIQMDKAAAFNOVLSQYH 2871  
QY 181 ---SD---L---L---A---S---G 187  
Db 2872 AELIILHTASDEYTEALEMFKNLIPINYQSFKPPhRAAICRISPSRVLCOVENSHAITDG 2931  
QY 188 ---S---G---DST---AR---I-W--- 196  
Db 2932 ASTSILANDLLOAYNGNSMPLNMD-TACEFARAQLTSSFGKLSYWKKKLREMDPCHFP 2990  
QY 197 ---NLS-E-----N-ST-- 203  
Db 2991 KISGASTQGTGTSVCKIRGALFASKIODYCN-SVEVTTASLFTQIWTALTAAYTGNOSTCF 3049  
QY 204 ---SG---S---T---QLV---L-RHC--- 214  
Db 3050 GYLASGRDLPIAGIDKSIGFTNMLVCRVINRETEILOFVQIVDQVMQDLEHQCSLA 3109  
QY 215 -IR-----EG-----QDVPS-----N--- 225  
Db 3110 SIQHELGINDNPLFNSILSYQKQDDEPAGDEGLVIKALDQD-PTEYDIVLNGHATDH 3168  
QY 226 ---K-----D-----V-T--- 229  
Db 3169 IEIVFDYKACLSSIOAESVLSMQSTAAALVOHASGDHQTLSRVNVMSTEDISDIWQWN 3228  
QY 230 S-----L-D----- 232  
Db 3229 SDVPVTVDCCVHHIITCHKRPQAPAI CAWDGDTVAEYNKLSDKLAHLVSVGVGPGV 3288  
QY 233 ---W-----N----- 234  
Db 3289 WVPLCEKSKWTPITAMVMKAGASVAMDSTQPEERLRAIVNQVKSPIILSSPFANEQLA 3348  
QY 235 ---SE---G-----TLJ-----A-T-- 242  
Db 3349 SRLISELPAHQHPNKQRSGKFCSEWKPLPHVNPDSITLVVFTSGTGVKGVAVTHS 3408  
QY 243 ---G---SY-D---G-----F-- 248  
Db 3409 NIASAIKHORHLGFTSESRVDFSSYFVWVWNCNLLQGLSAGSCVCI PSDNERKTDfMA 3468  
QY 249 ---A----- 249  
Db 3469 AIVKRNANLVITPSAIRGLKLDALNSLCNVHFI GEPLHVDTFRSVDESVTISNLYGTE 3528  
QY 250 ---R---I---W---T----- 253  
Db 3529 CTTFTSTVOTICGROHQHSITIGKAGLNTWADIATGALVPIGSAGELLLEGPLVAAGYR 3588  
QY 254 ---K---D-----G-----NLAS--TL-- 262  
Db 3589 GDAVKTAATAFYVDPFLLRGVGHGRRGRGLYKTGDIVRYN--SNGLTFLGRKDSQVKI 3646  
QY 263 -GO-----H-----K-----GP----- 268  
Db 3647 NGQVERFEDIESHINGALLPDFSEGOALVDFVTPQSSRPMLVAFVFGPTVTTEGMDDEAD 3706

QY 269 ---I-F-----A-----L-K--WN-- 275  
Db 3707 ILSLAKRTAISLDESIAARI PAFMIPSAVIPIQKIPVTATGKTRRRRLREMAKDVTDQL 3766  
QY 276 -K-----K-G-NP----- 280  
Db 3767 IKADSHGPDRCQGTMEIQLQILMGTVLGVESSLIGAHDNFMRVGGDSVGAI RLASSAR 3826  
QY 281 ---IL-----SA--GV 286  
Db 3827 ELGFTLNADILKNPKLSDMAKIMIRTPSQDISIKFSLLPKPGSDVNWAVAE TSALCGV 3886  
QY 287 D-----K-----T----- 290  
Db 3887 DGNQVEDLYPCTPQEGLLATTKPGDYIIRCILELKSTDVKKFCASWEAVELESTPIL 3946  
QY 291 ---I---W---D---A----- 295  
Db 3947 RTRIVDIAEQGLVQAVIKQPAQWTSAEASSLVDFVAADNEKTTLGLMPLVRFGLVQETNK 4006  
QY 296 H---T-----G-E---AK--- 301  
Db 4007 HFFVLTHHAVVDGWNALNVFEKLENFYAGSRHESPDPRHFVKHISLNDNDAAAFWKD 4066  
QY 302 Q-Q-----FP-----H----- 307  
Db 4067 QLOGSEAPTFPSLPTATFVPKSEKTLIHTVEELQWPKTNVTAFTLVRAALSLLTAAYTNS 4126  
QY 308 ---S-----A-P-----A-L---D----- 313  
Db 4127 EDVCFGTSGRQVGLPGVVERMIGPTIATVPVRVIDREQLQAFLTOMQHOSIDMIAFE 4186  
QY 314 ---V---D-----WOS----- 318  
Db 4187 QFGLQOIRKSSPDARACNFQSLIIVQPAEETAQWOSDIIARDIGEGADDPMGIOBIGTY 4246  
QY 319 ---N-N-----T-F---ASC-----S---T-- 327  
Db 4247 ALTLECHLGPDSLLIKANFDSNVIDELOVKRFTKQFEHVLRLQICCSGSLVWSDIDTTSR 4306  
QY 328 -DMC-----I---H---VC----- 334  
Db 4307 QDMEDIKWNNAVPOSVNTFVHELISVARRLPHVQAVCAWGNWNTYRQLDLSNVVAHH 4366  
QY 335 ---K-----K-----LG-----Q-I- 342  
Db 4367 LVDLGVGSQDIVPLLFKEKSWPIAMLGVMKAGASA VADTSQPKDRLRMIIDQANPTVA 4426  
QY 343 ---K---T---F---QG----- 347  
Db 4427 LSSADKLPLVRSITKAQSFVWSQGDRLKPSLNATLPVWDPSRLYLVTSGSTGVPK 4486  
QY 348 ---H-----T----- 349  
Db 4487 GVIIHRCNFASAIKHOKEVQGLPSTRVYDFASYAFDVANALLTFESGACLICPSDAD 4546  
QY 350 ---N-----E-----V- 352  
Db 4547 RKNDLNGSTARLKPHTADLTPSAALVLSKESLOQDRTLGLGERLLAEYATKWSQFVTVK 4606  
QY 353 N-----AI-----K-----W-DP-TG----- 361  
Db 4607 NSYGPSECTPTATFTEAIGRGYDLGASIGK PAGLNTWVDPVTGQSPLVPIGGVGLFLEG 4666  
QY 362 ---N-----LL----- 364  
Db 4667 PLVGAGYLDAAEKTAAFIHDPPFLLRGNVVAQFGRRTLYKTGDIVRNSDGLTFVWR 4726  
QY 365 ---A---S---C-----D--- 370  
Db 4727 KDTQVKINGORVELABIESHIALYATROVATLPTGLCANKLVAMISLTDVNDYVSED 4786  
QY 371 M-----TL---KI---W-----SM---KQDN--- 382

Db	4787	: :	LAENKIELASSEHDOLINEHIEALQSLRLRESLPQYMIPSLVWVLNLYNLPMTASGQDNKAL	4846
Qy	383	-	-	383
Db	4847	KSWLENWDETLPFSKINNANGSDIIRKPDTEDERVLSQKCSIVLNMPPVDKINLKDSFIANG	4906	
Qy	384	-	-	387
Db	4907	GDSISAMRLASHYRTVGISVSITLLQSOKTLADPAAFSGATAISGVSQEHTDVPFELSP	4966	
Qy	388	--OO-H--NKE-IY-T-I-	-	397
Db	4967	IQQWFDPQSPMQOQHDFYN-QGFYVRILRTRVINDLES AFLSVLRHAMLRSRFQH	5025	
Qy	398	--KW--S-	-P-	401
Db	5026	GGKKWKIILSHSKRALHLNVSOHLSMEIASLAQERHQIDIEKGPVFSDICLLGOQQH	5085	
Qy	402	--T--G-	-P-	404
Db	5086	LVMIAHLVTDLVSWRIILDLETILNGHSITAALPFQVWSRLQAERAVSTLKPNNLS	5145	
Qy	405	--GT-NY-:PN-AN-LML---	415	
Db	5146	TGCVHNKLKFVKYTHDTPNCPLADHLRSVTIDRETTAVLLGEANNMTPPEVILLASVV	5205	
Qy	416	-A-SA-S-F-P-D-ST-VRL	426	
Db	5206	DAPFTFSQRNSLTIFNEGHGREAWSDEILDSTVGWFTTLSPINIYRNATSETDMVRL	5265	
Qy	427	--W-	-DVD	430
Db	5266	VKDARRSLPANGMSYFTSRYLNPDGORAPESHNTVSEVFNHYHCQFOQLSHQALPEDID	5325	
Qy	431	--R-G-	-IC-	434
Db	5326	LVGVRVQGRSIGASGLNFIEVAIEAMQAHFEFSVNQNIAHQSLTNOWIDIQPSLERICL	5385	
Qy	435	I-HTL-   T-	-	439
Db	5386	VLEANTHTLCDPKFISLDYQRLDDLTSLRPBESINSTVEEPSCSPIVDGMLLSQ	5445	
Qy	440	-K--H--Q-EP-V-Y--	446	
Db	5446	IKQPESYKTILQRYEVLSHSDHPICDLTKIAWQRVISRQPALRTVFIAGLDGSTAFYQAL	5505	
Qy	447	-S-V-----APS-	-PD-	453
Db	5506	LKCQSGDIVVEAKTEBEALKAFSSLPKVYQQAQPPHRLTLCQTDPDKVCQIEMSHAI	5565	
Qy	454	-G----	-G-	454
Db	5566	TDGASTILIKOLIDAYGDRLSTDVLKTTTFEFASHLLAKPQSQKISYWNTKLGLEPCR	5625	
Qy	455	-R-	-	455
Db	5626	FPSLSSMSRKHECSSSIGVFVEDKMPAQIQDFPCSNQVTPASLLKSAAWALTISTYVQNO	5685	
Qy	456	-YLASG-	-	460
Db	5686	SVCFGYLASGRDLPIAGMDSEVGAYTNIMVCRADLDGQPGVALVRQLNQLMQDLSFQH	5745	
Qy	461	-SF---D-K-	-	464
Db	5746	ISLASIQHELGLASDOOLFNSIVSFORSQDDNEQSAPEGKLRPFNINDGLDPTBYDIVLGI	5805	
Qy	465	-C- -V-H- -WN-T-471-	-	471
Db	5806	NQOTRSTIEDLFSSHCLTSNQAKRILEHLOSNIAAILHNEPPALISPQEQDIWSNST	5865	
Qy	472	QV-CLH- -Y- -L-N-	-	479

Db	5866	-VPMNVCVHDHLSIKVIFRQPDAPAVCSWDGDFTFVAELDNLATRLANLSKMGIGRGS	5924
Qy	480	-----G-----Q-----	481
Db	5925	VPLCFEKSQWTPYVLAAMVKMTGAASVTMDTSQPEERLQSIQAQVDAKLVISSTLKVELAA	5984
Qy	482	-----VL-----L-----NL-----	486
Db	5985	RLTTAPVLAIDKASKMKAMADDTPLAAVDPAANSIYIVFTSGSTGTPKGVIIHTHTNYS	6044
Qy	487	-----G-----R-----S-----I-CL-----	492
Db	6045	HQOSEHGFKPSTRVDFDPASYAFDVSWSNFLHTLTIGACLCIPSDHDKNDPAGAI	6104
Qy	493	-----Y-----	493
Db	6105	THVDMTPSAASVLPASTLAKLDTIVLGEKLSLEYAQRWSALTSVRNPYPGPSECTPT	6164
Qy	494	T-----L-----P-----H-----	499
Db	6165	TEINSABISKGVSKIGKVGGLNTWIVDPATAQHLMPGIPGELLLEGLVAGYLGDPVK	6234
Qy	498	-----	497
Db	6225	TASAFIEDPFLVKGASPGIPGRGRRLYRTGDLVTYNTDGSLSFVGRDRSQIKNGOR	6284
Qy	498	-----H-----LV-----VI-P-----LVA-----L-Y-----E-LL-----	511
Db	6285	LGDIESHVSANLVSHGSAQVAVEVSPQASSNNILVAFVFDLNSLNINDEKLARTKA	6344
Qy	512	-----VL-----K 514	
Db	6345	ATEGIREKLATQIPSYMIPSWIPTVTFPTATGK 6379	
RESULT 13			
ABP60991			
XX	ID	ABP60991 standard; protein; 5635 AA.	
XX	AC	ABP60991;	
XX	DT	10-SEP-2002 (first entry)	
XX	DE	Novel human protein. SEQ ID 78.	
XX	XX	Human; cytostatic; vulnery; antiarteriosclerotic; antiparkinsonian;	
KW	KW	neotropic; neuroprotective; immunosuppressive; haemostatic;	
KW	KW	antiinflammatory; cardiact; antitumor; virucide; antithyroid;	
KW	KW	cerebroprotective; anorectic; metabolic; vaccine; cancer; infection;	
KW	KW	wound healing disorders; atherosclerosis; parkinson's disease;	
KW	KW	Alzheimer's disease; autoimmune disorder; haematopoietic disorder;	
KW	KW	inflammation; neoplastic disease; nervous system disorder;	
KW	KW	cardiovascular disorders; pancreatitis; respiratory disorder;	
KW	KW	hyperproliferation; systemic autoimmune disease; hyper-immunity;	
KW	KW	developmental abnormality; gastrointestinal ulceration; neuropathy;	
KW	KW	haematological disease; metabolic disease; sperm dysfunction;	
KW	KW	thyroid disorder; hypothyroidism; brain damage; colitis;	
KW	KW	cone photo- transduction deficiency; neurological disease; stroke;	
KW	KW	angiogenesis; ovulation disorder; spinal cord; thyroid gland; heart;	
KW	KW	trachea; thymus; lymph node; muscular system; obesity; anorexia;	
KW	KW	growth abnormality; precocious puberty.	
OS	OS	homo sapiens.	
XX	XX		
PN	PN	W0200250105-A1.	
XX	XX	27-JUN-2002.	
XX	XX	17-DEC-2001; 2001W0-US049232.	
PF	PF		
XX	XX	19-DEC-2000; 2000US-0256710P.	
PR	PR	20-DEC-2000; 2000US-0257048P.	
PR	PR	09-JAN-2001; 2001US-0260482P.	

PR 30-JAN-2001; 2001US-0264922P.  
 PR 06-FEB-2001; 2001US-0266797P.  
 PR 19-MAR-2001; 2001US-0276988P.  
 PR 04-APR-2001; 2001US-0281535P.  
 PR 08-MAY-2001; 2001US-0289622P.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM CORP.  
 PA (SMIK ) SMITHKLINE BEECHAM PLC.  
 PA (GLAX ) GLAXO GROUP LTD.  
 XX  
 PI Agarwal P, Birkeland M, Cogswell JP, Kabnick KF, Lai Y;  
 PI Martensen SA, Rizvi SK, Smith RF, Strum JC, Xie Q;  
 XX  
 DR WPI; 2002-508784/54.  
 DR N-PSDB; ABQ86156.  
 XX  
 PT Secreted proteins and polynucleotides useful as vaccines for preventing  
 PT or treating various diseases e.g. cancer, wounds, atherosclerosis,  
 PT Parkinson's disease, Alzheimer's disease, infection, autoimmune disorder.  
 XX  
 PS Claim 1(a); Page 285-297; 335pp; English.  
 XX  
 CC The invention relates to an isolated polypeptide with signal sequences  
 CC which allow it to be secreted extracellularly or membrane associated. The  
 CC activity of polypeptides of the invention may be described as,  
 CC cytostatic, vulnerary, antiarteriosclerotic, antiparkinsonian, nootropic,  
 CC neuroprotective, immunosuppressive, haemostatic, antiinflammatory,  
 CC cardiant, antitumor, virucide, antithyroid, cerebroprotective, anorectic,  
 CC and metabolic. Polypeptides and polynucleotides of the invention are  
 CC useful in the treatment, or as a vaccine in the prevention of, cancer,  
 CC wound healing disorders, infection, atherosclerosis, Parkinson's disease  
 CC and Alzheimer's disease, autoimmune disorder, haematopoietic disorder,  
 CC inflammation, neoplastic diseases, nervous system related disorders and  
 CC cardiovascular disorders, pancreatitis, respiratory disorder,  
 CC hyperproliferation, systemic autoimmune disease, hyper-immunity,  
 CC developmental abnormality, gastrointestinal ulceration, neuropathy,  
 CC haematological diseases, metabolic diseases, sperm dysfunction, thyroid  
 CC disorders e.g. hypothyroidism, brain damages, colitis, cone photo-  
 CC transduction deficiency, neurological diseases, stroke, angiogenesis,  
 CC ovulation disorders, diseases in the spinal cord, thyroid gland, heart,  
 CC trachea, thymus, lymph node and muscular system, obesity, anorexia,  
 CC growth abnormalities, and alleviation of precocious puberty. The  
 CC sequences given in records ABP60965-ABP61019 represent novel human  
 CC proteins of the invention  
 XX  
 SQ Sequence 5635 AA;

Query Match 71.3%; Score 2601.6; DB 5; Length 5635;  
 Best Local Similarity 9.8%; Pred. No. 8.5e-31;  
 Matches 431; Conservative 67; Mismatches 13; Indels 3904; Gaps 372;  
 1 M-S-----I-----SS---D-----EV-----8  
 1 MISWEVVHVFLLFALLYSLAQPASQSEIRABEIEGASTLAFVFDVTGSMYDDLQVVI 60  
 9 -----NF-LV-----12  
 61 EGASKILETSLEKPEPLNFALVPDPHDEIGPVTTITDPKKFQYELRELYVQGGDCPE 120  
 13 -----Y-----R---Y-----LQ-----17  
 121 MSIGAIAKIALEISLPGSFYVFTDARS KDYRLTHEVLQIQKQSQVFWLTGDCDDRTH 180  
 18 -----E-----SG--F-----S-H--S-----A-----F 26  
 181 IGKVVETIASTSGQVFLDKQVNEVLKWTVEAVQASKVHLLSTHLEQAVNTWRIFP 240  
 27 -----T-----FG-----I-----K 31  
 241 DPSLKEVTVSLGSPSPMIEIRNPLGLIKKGFGLHELLHNSAKVNVYKEPEAGWTVK 300  
 32 ---S-H-----I-----S---Q-----36

Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFKKTVSRRPVQGIPTVYLLNTSGISTPARID 360  
 QY 37 -----S-----NI-----N-----G-----41  
 Db 361 LLELLSISGSSSLKTIPIVKYPPHRKPYGIWNISDFVPPNEAFFLKVTGYDKDKDYLFORVSS 420  
 QY 42 -----A-----L---VP-----P-----A 47  
 Db 421 VFSISIVDPAPKVTMPKTPGYYLQPGQIPCSVDSLLPFTLSFVRNGVTIGVDQYKESA 480  
 QY 48 -----A---L-----I-----S---IIQ-----KG-----56  
 Db 481 SVNLDIAKVTLSDEGFEICIAVSSAGTGRAQTFDFVSEPPVIVQVNNVTVTGERAVLT 540  
 QY 57 -----L-Q-----Y---V---E-----61  
 Db 541 CLIIASVDYNTQWRNDRVRLAEPARTLANLSLELKVRFNDAGEYHCVMSSEGGSS 600  
 QY 62 -A-----EVS-----NE-----68  
 Db 601 AASVFLTVQEPKVTVMKQNSFTGSGSEVSMCSATGYPKPKIAWTVNDMFIVGSHRYRM 660  
 QY 69 --DGLTF-----D-----G-----R---P-----I--78  
 Db 661 TSDGTLFIKNAAPKADAGIYGCLASNSAGTDKQNSTLRYIEAPKLMVVSSELLVALGDITV 720  
 QY 79 -E--S-----L-S--LI-----DA-----V-----87  
 Db 721 MECKTSGIPPPQVKNFKGDLRLPSTFLIIDPLGLLKIQETQDLDAQDYTCVAINAAGR 780  
 QY 88 -----M-----P-DV-----VQ-----T-----94  
 Db 781 ATGKITLDVGSPPVFTQEPADYSMEIGSNVILPCYVQGYPEPTIKWRRLDNMPISRPSPS 840  
 QY 95 -----R-----95  
 Db 841 VSSISQRTGALFILNLWASDKGTVICEAENQFKIQSETTVTGTGLVAPLIGISPSVAN 900  
 QY 96 ---QQ-----A-----Y-----R---D--KL--103  
 Db 901 VIEGQQLTPTLLAGNPPIPERWIKNSAMLLQNPIYTVRSGLSHIERVQLQDGEYTC 960  
 QY 104 -----A-----Q-QQ-----AA-----109  
 Db 961 VASNVAGTNKTTSVVHVLPITQHQQLLSTIEGIPVTLPCCKASGNPKPSVWSKKGEL 1020  
 QY 110 ---AAA---A-A-----A-AA-A-----118  
 Db 1021 ISTSSAKFSAGADGSLYVSPGGEESGEYVCTATNTAGYAKRKVQLTVYVRPVFGDQRG 1080  
 QY 119 -SQ-----Q-----Q-----GS-----123  
 Db 1081 LSQDKPVEISVLAGEVTLPCVKSLPPIITWAKETQLISPFSPRHTFLPSGSMKITET 1140  
 QY 124 -----A-----A-----K-----N-----126  
 Db 1141 RTSDSGMYLVATNAGNVQAVKLVNVHPVKIQRPKHLKVQVQGVQVVDIPCNAQGTPLP 1200  
 QY 127 -----GE---N-----T-A-N-G-----133  
 Db 1201 VITWSKGGSTMLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260  
 QY 134 -----E---N-----CA-----H-----139  
 Db 1261 PPTVEDLEPPYNTTQERVANQRIEPPCPAKTPKPTIKWLNHRELGTREPGISILEDG 1320  
 QY 140 T---IA-----N-----HT---DM-----148  
 Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHPPVPIKDKQVTVNSVLLNLQNL 1380  
 QY 149 -----M-----E-----150  
 Db 1381 FCEVEGTSPPIIMWYKDNVQVTSSTIQTVNNKGILKLFRTATPDAGRYSCAKINIAIGTS 1440

QY 151 ---VD---G---DV-E---I---P---PN 160  
Db 1441 QKYNIDVLVPPTIIGTFNPEVSVLNRDVALEQVKGTFFPDIIHWFKDGKFLFLGDPN 1500  
QY 161 ---K-A---V---VL 165  
Db 1501 VELLDRGQVLHLKARNDKGRYQCTVSNAAGKQAKDIKLIYNPSPISKGGNVTTDISVL 1560  
QY 166 ---RG---H- 168  
Db 1561 INSILKECTRLPMPAITWYKQOPIMSSQALYIDKQYLHIIPRAQVSDSATYTCV 1620  
QY 169 ---E-S---E---VF-I---C-A---W--- 177  
Db 1621 ANVAGTAESFHVVDVVPVMEIGNLATPLNKQVIAHSLTLECKAAGNPSILTLKDG 1680  
QY 178 ---N---P-V-SD- 182  
Db 1681 PVKANDNIRIEAGKKLEIMSAOEIDRGQYICVATSVAGEKEIKEYVDVLVPPAIEGGDE 1740  
QY 183 ---LL---ASGS---G---D---S--- 191  
Db 1741 TSYFIVWNNLELCHVTGSPPTIMWLKDGQIDRDERGFKILNGRLVIAQVSN 1800  
QY 192 ---TA---R---I-W 196  
Db 1801 GLYRCMAANTAGDHKEFEVHVHPPTIKSSGLSERVVVKYPVALQCIANGIPNPSITW 1860  
QY 197 ---NL---T---SGST---S 199  
Db 1861 LKDDQPVNTAQGNLKTQSSGRVLIQIAKLTLEDAGRYTCVATNAAGTQHQIQLHVHPPS 1920  
QY 200 -E---N---S---T---SGST--- 207  
Db 1921 LEDAGKMLNETVLVSNPVOLEKCAAGNPFVITWYKDNRLSGSTWTFNLRGQIIDIES 1980  
QY 208 -Q---L-V---LR-H---C---I- 215  
Db 1981 AQISDAGIKYCAINSAGATELFYSLOVHVAPISGSNNVAVVNNPVRLECEARGIPA 2040  
QY 216 ---R---E-G---G-Q---D--- 221  
Db 2041 PSLTMLKDGSPVSSFNGLQVLSGGRILALTSAQISDTGRYTCVAVNAAGEKQORDIDLRV 2100  
QY 222 -VPSN---KD---VT---S-L- 231  
Db 2101 YVPPNIMGBEQNVSLISQAVELLQSDAIPPTTLTWLKDGHPLLLKPKGLSISENRSVLK 2160  
QY 232 ---D---W-N---S---EG---TLL-A 241  
Db 2161 IEDAQVODTGRYTCATNVAGTEKYNVNIWVPPNIGSGDELTLQTLVIEGNLISLLES 2220  
QY 242 T---GS-Y-D---G---F---A- 249  
Db 2221 SGIPPNLIWKKKSPVLTDSMRVRLSGRQLQISIAEKSDAALYSCVASNAGTAKK 2280  
QY 250 ---R---I---WTKDG--- 256  
Db 2281 EYNLQVYIRPTITNSGSHPTIIVTRGKSISECEVOGIPPTVTWMDKGLHLIKAGVE 2340  
QY 257 ---NL-AS---T---L---GOH--- 265  
Db 2341 ILDEGHILQKNIHVSDDTGRYCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVVE 2400  
QY 266 K---G-P-I-FALK-W--- 274  
Db 2401 KNSVSLTCEASGILPLPSITWF-KDGPVLSNSVRILSGGRMLRLMQTTMEDAGQYTCV 2458  
QY 275 ---N---K-K---GN--- 279  
Db 2459 VRNAAGEERKIFGLSVLVPPIHVIGNTLBDVVKKEKQSVTLTCEVTGNPVPBITWHKQOQ 2518

QY 280 ---F-I---L---S---AGV- 286  
Db 2519 PLQDEAHHIIISGRFLQITNVQVPHTRGYTCLCLASSPAGHKSRFSLSLNVFVSPTIAGVGS 2578  
QY 287 D---K-T---T-I---I-W--- 293  
Db 2579 DGNPEDVTILNSPTSLVCEAYSYPATITWFKDGTLESNNRIRILPGGRTILQILNAQE 2638  
QY 294 D---A-H---TG---EA 300  
Db 2639 DNAGRYSCVATNEAGEMIKHYEVKVYIPPIINKGDLWGFGLSKPEVKIKVNNLTLECEA 2698  
QY 301 ---K---Q---Q---F 304  
Db 2699 YAIPASLSWYKDGQPLKSDDDHVNIAANGHTLIQKEAQISDTGRYTCVASIAGEDELDF 2758  
QY 305 ---P-F---LDV---H- 307  
Db 2759 DVNIQVPPSFQKLWEIGNMLDTRNGEAKDVIINNPFISLYCETNAAPPPTLTWYKDGHL 2818  
QY 308 -SA---P---A---LDV---D- 315  
Db 2819 TSSDKVLIILPGRVLOIPRAKVEDAGRYTCVAVNEAGESLQYDVRVLVPPPIIKGANSDL 2878  
QY 316 ---WQ---SN--- 319  
Db 2879 PEEVTVLVNKSALIECLSSGSPAPRNSWKDGQPLLEDDHKKFLSNRILQILNTQITDI 2938  
QY 320 ---NT---PAS-C---S- 326  
Db 2939 GRVYCAENTAGSAKKYFNLNVHVPSPVIGPKSENLTVVVNNFISLTCEVSGPPDLSW 2998  
QY 327 ---T---D---MCI--- 331  
Db 2999 LKNEQPIKLTNTLIVPGGRTILQIRAKVSDGGEYTCIAINOAGESKKPSLTVYVPPSI 3058  
QY 332 -H---V---C---K---L-GQ--- 338  
Db 3059 KDHSLSLVNVREGTSVLECESNAVPPVITWYKNGRMITESTHVELADQMLHIK 3118  
QY 339 ---D---R---P-IK---TF--- 345  
Db 3119 KAEVSDTGOYVCRAINAGRDDKNFHLNVYVPPSIEPEREVIVETISNPVTLTCDATGI 3178  
QY 346 ---Q---G---H---T--- 349  
Db 3179 PPTTIANLKNHKRIENSLSLEVRILSGSKLQIARSQSDSGNYTCIASNMEGAOKYF 3238  
QY 350 ---N-E-V-NA---IKW-D-P---T--- 360  
Db 3239 LSIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIQLMKDKGPIASGETERIVS 3298  
QY 361 -G---NLL-A-S---C-S---D---D---MTL--- 373  
Db 3299 ANGSTLIYALTSDTKYTCVATNPAGEBDRIFNLNVYVPTIRGNKDEAEKMLTLDVT 3358  
QY 374 ---K---I-W---SM---KQ-D-D---NCV--- 384  
Db 3359 SINIECRATGTPPPQINLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSVAVVTCVASNR 3418  
QY 385 ---H---D---L---Q--- 388  
Db 3419 AGVDKNHYNLQVFPAPNMDNSMGTBEITVLKGSSTSMACITDGTAPSMANLWLDGQPLGL 3478  
QY 389 ---Q---H-N---KEI 394  
Db 3479 DAHLTVSTHGMVQLLKAETEDSGKYTCIASNEAGEVSKHFKVLEPPHINGSSEHEEI 3538  
QY 395 ---YT---I---K-W--- 399  
Db 3539 SVIVNPLELTCTIASGIPAPKMTWMKDGRLPQTDQVOTLGGGEVLRISTAQVEDTGRYT 3598  
QY 400 ---SPTG---P---GT---N---N- 408

```

Db 3599 CLASSPAGDDKEYLVVRVHPNPNTAGTDEPRDITVLNRQVLTLECKSDAVPPPVITWLRN 3658
Qy 409 -----P-----N-----ANLM-LAS--A---S--F----- 420
Db 3659 GERLQATPRVRLSGRYLQINNDALDGTANTYTCVAGNIAGKTTREFILATVNVPPNIKGG 3718
Qy 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433
Db 3719 POSLIVILLNKSTVLECTAEGVPTTRITWRKDGAVLAGNHARYSILENGFLHIOSAHVTD 3778
Qy 434 -----C-----I-----H-----TL----- 439
Db 3779 GRYLCMATNAAGTDRRIDLQVHPPSIAPGPTMTVIVNVQITLACEATGPKPSINWR 3838
Qy 440 K--H-----Q-E-----P-V-----Y-----S 447
Db 3839 KNGHLNVNDQNSRYLLSSGSLVILISFSDDTATYECTVTNGAGDDKRTVDTLVQVPPS 3898
Qy 448 VA-----F-S-----P-D----- 453
Db 3899 IADEPTDLTKHAPAVITCTASGVPPPSIHWTNGIRLLPRGDGYRLSSGAIEILATQ 3958
Qy 454 -----GRY-----L---ASG--S-F 462
Db 3959 LNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIOQPSELHVILNPNILLPCEATGTPSPF 4018
Qy 463 -----L--G-----D-----KCV----- 466
Db 4019 ITWQKGINVNTSGRNHVLPSGLQISRAVRDAGTMCVAQNPAQTALGKILNVQVP 4078
Qy 467 -----H-----I-W-----NT-Q--V 473
Db 4079 PVISPHLKEVIAVDKEITLSCADGLPPDITWHKDGRAIVESIRORVSSGSLQIAFV 4138
Qy 474 -----C-----L-----HY----- 477
Db 4139 QPGDAGHYTCWAANVAGSSSTSKLTIVHPPIRSTEGHYTVNENSOAILPCVADGIPTP 4198
Qy 478 -----L-N-----GO-----VL-----L-----N----- 485
Db 4199 AINWKKDVLNALLGKYTAEPYGELENVVLEDSGFYTCVANNAAGEDTHTVSLTVHV 4258
Qy 486 -----L-G--R-SIC-----L-VT-----LP-H-----H--LV--- 500
Db 4259 LPTFTLPGDVLNKGELRLS-CKATGIPLPKLTWTFFNNIIPAHFDSVNGHSELVIER 4317
Qy 501 -----V--I-----P-----L-----V-----A-L----- 507
Db 4318 VSKEDSGTYVCTAENSVGFKVKAIGFVVVKEPPVFKGDYPSNWIEPLGGNAILNCEVKGDP 4377
Qy 508 -----IE-----L-----L-VL-----K 514
Db 4378 TPTIQNRKGVDEISHRIQLNGSLAIYGTVNE 4412

```

## RESULT 14

ADJ70089

ID ADJ70089 standard; protein; 5636 AA.

XX AC

XX AC

XX ADJ70089;

DT 06-MAY-2004 (first entry)

DE Human heat mitochondrial protein as a therapeutic target SeqID1895.

XX KW

KW mitochondrial; human; screening assay; diabetes mellitus;

KW Huntington's disease; osteoarthritis;

KW Leber's hereditary optic neuropathy; LHON;

KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;

KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;

KW neuroprotective; nontropic; antidiabetic; anticonvulsant; antiarthritic;

KW osteopathic; ophthalmological; cytostatic.

XX OS Homo sapiens.

XX PN WO2003087768-A2.

XX PD 23-OCT-2003.

XX PF 04-APR-2003; 2003WO-US010870.

XX PR 12-APR-2002; 2002US-0372843P.

XX PR 17-JUN-2002; 2002US-0389987P.

XX PR 20-SEP-2002; 2002US-0412418P.

XX PA (MITO-) MITOKOR.

XX PA (BUCK-) BUCK INST AGE RES.

XX PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;

XX PI Warnock DE;

XX PI WPI; 2003-845369/78.

XX DR

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

XX CC

Query Match 71.2%; Score 2596.5; DB 7; Length 5636;

Best Local Similarity 9.7%; Pred. No. 1.1e-30;

Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

Qy 1 M-S-----I-----SS--D-----EV----- 8

Db 1 MISWEVHTVFLFALLYSLAQDASFSQSEIRAEPEPEGASTLAFVFDVTGSMYDDLQVVI 60

Qy 9 -----NF-LV----- 12

Db 61 EGASKILETSLRKPXPLENFALVPHPDPEIGPVITTPDKKFOYELRELYVQGGDCPE 120

Qy 13 -----Y-----R---Y-----LQ----- 17

Db 121 MSGAIIKIALEISLPGSFYVFTDARKSDYRLTHEVLQIQKQSOVFLVLTGDCDDRTH 180

Qy 18 -----E-----SG--F-----S--H--S-----A-----F 26

Db 181 IGYKVEEIASTSSGQVFLDKKQVNEVLKWVEAVQASKVHLLSTDHLEQAVNTWRIFP 240

Qy 27 -----T-----PG-----I-----K 31

Db 241 DPSLKEVTVSLSGSPMIEIRNPLGKIKGFGHLHLLNHSKVVNVKEPEAGNWTVK 300

Qy 32 -----S--H-----I-----S--Q----- 36

Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDPKKTVSRPVQGIPTVYLLNTSGISTPARID 360

QY 37 -----S-----NI-----N-----G----- 41  
Db 361 LLELLSISGSSLKTIPVKYYPHRKPYGIWNISDFVPNEAFFLKVTGYDKDDYLFORVSS 420  
QY 42 -----A-----L-----VP-----P-----A 47  
Db 421 VSFSSIVDPAPKVTMEKPTPGYYLQPGQIPCSVDLSLLPFTLSFVRNGVTGLQDQYLKESA 480  
QY 48 -----A-----L-----I-----S-----IIQ-----KG----- 56  
Db 481 SVSLDIKVTLSDEGFYECIAVSSAGTGRAQTFDFVSEPPPVIOVNNVTVPGERAVLT 540  
QY 57 -----L-----Q-----G-----R-----P-----Y-----V-----E----- 61  
Db 541 CLIIISAVDNLTWQRNDRVRLAEPARIRTLANLSLELKSVEKFNDAAGEYHCWSSSEGGSS 600  
QY 62 -A-----EVSII-----NE----- 68  
Db 601 AASVFLTVQEPKVTYMPKNQSFSGSEVSIMCSATGYPKPKIAMTWNDFIVGSHRYRM 660  
QY 69 -----D-----G-----R-----P-----I----- 78  
Db 661 TSDGTLFIKNAAPKADAGIYGCLASAGTDKQNSTURYIAPKLMVQSELLVALGDIIV 720  
QY 79 -E-----S-----L-----LI-----DA-----V----- 87  
Db 721 MECKTSGIPPPQVKFKGDLRLPSTFLIDPLGLLKIOETQDLADGYTCVAINENGR 780  
QY 88 -----M-----P-----DV-----VO-----T----- 94  
Db 781 ATKITLGVGSPVFTQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRRLDNMPIFSRPFS 840  
QY 95 -----R----- 95  
Db 841 VSSISQRTGAULFILNLWASDKGTIYCEAENQPKIQSETTVTVTGLVAPLIGISPSVAN 900  
QY 96 -----QO-----A-----Y-----R-----D-----KL----- 103  
Db 901 VIEGQQLTLPCTLLAGNPPIERWIKNSAMLLQNPIYTVRSDSGLHIERVOLQDGEYTC 960  
QY 104 -----A-----QO-----AA----- 109  
Db 961 VASNAVGTNNKTSVVVHVHPTIHQQLSTIEGIPVTLPCKASGNPKPSVIWSKKGEL 1020  
QY 110 ---AAA---A-A-----A-AA-A----- 118  
Db 1021 ISTSAKFSAGADGSLYVVSPEGESYVCTATNTAGYAKRKVQLTVYVRPRVFGDLRG 1080  
QY 119 -SQ-----Q-----GS----- 123  
Db 1081 LSQDKPVEISVLAGEBVTLPCEVKSILPPIITWAKETQLISPPSPRHTLPSPGSMKITET 1140  
QY 124 -----A-----K-----N----- 126  
Db 1141 RTSDSGMVLCVATNIAGNVTAQVKNLVHVPKPIQIRGPKHLKVQVGQORVIDPCNAOQTPLP 1200  
QY 127 -----GE-----N-----T-A-N-G----- 133  
Db 1201 VITWSKGGSTMLVDGEHVSNDPGTSLIDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260  
QY 134 -----E-----N-----GA-----H----- 139  
Db 1261 PPTVEDELPYNTTFOERVANQRIEFPCKAGTPKPTIKWLHNGRELTCREPGISILEEG 1320  
QY 140 T---IA-----N-----N-----HT-----DM----- 148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVPIKQEVSNVSVLLNQLTNL 1380  
QY 149 -----M-----E----- 150  
Db 1381 FCEVEGTSPPIIMWYKDNQVTSSTIQTVNNKGILKLFATPEDAGRYSCAKAINIAGTS 1440

QY 151 -----VD-----G-----DV-----E-----I-----P-----PN 160  
Db 1441 QKYFNIDVLVPPTIIGTNEPPEKVSVVLRNDRDALECOVKGTPTPPDIHWFKDGKGLFLGDPN 1500  
QY 161 -----K-A-----V----- 165  
Db 1501 VELLDRGQVLHLKNAARNDRKGRYQCTVSNAAGKQAKDIKLTIIYPPSIKGNVTDTISVL 1560  
QY 166 -----RG-----H----- 168  
Db 1561 INSLIKLECKTGLPMPAITWYKQOPIMSSSOALYIDKGOYLHI PRAQVSDSATYTCV 1620  
QY 169 -----E-S-----E-----VFI-----C-A-----W----- 177  
Db 1621 ANVAGTAEKSFHVVDVVPPIEMIEGNLATPLNKQVIAHSLTLECNAAAGNPSILTLKDG 1680  
QY 178 -----N-----P-V-SD- 182  
Db 1681 PVKANDNFRIEAGKKLEIMSAQIEDRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740  
QY 183 -----LL-----ASGS-----G-----D-----S----- 191  
Db 1741 TSYFIVWNNLLELDCHVTGSPPTIMLWKDQQLIDERDQPKILLNGRKLVIQAQVSNT 1800  
QY 192 -----TA-----R-----I-W 196  
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPTIKSSGLSERVVVKYPVALQCIANGIPNPSITW 1860  
QY 197 -----NL-----SGST-----S 199  
Db 1861 LKDDQPVNTAQNLIKIQSSGRVLQIAKTLELDAGRYTCVATNAAGETOQHQLHVHPPS 1920  
QY 200 -E-----N-----S-----T----- 207  
Db 1921 LEDAGKMLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLSGSTMTFLNRQIIDIES 1980  
QY 208 -Q-----L-V-----LR-H-----C-----I-- 215  
Db 1981 AQISDAGIVKCAINSAGATELFYSLQVHVAPSISGSNNMVAVVNVNVRLECEARGIPA 2040  
QY 216 -----R-----E-G-----G-O-D----- 221  
Db 2041 PSLTWLKDGPVSSFSNGLQVLSGGRILALTSTQISDTSRYTCVAVNAAGEKORDILRV 2100  
QY 222 -VPSN-----KD-----VT-----S-L- 231  
Db 2101 YVPPNIMGEQNVSVLISOAVELLQSDAIPPTTLTWLKDGHPLKKPGLSISENSVLK 2160  
QY 232 -----D-----W-N-S-----EG---TLL-- 240  
Db 2161 IEDAQVQDTRVTCATNVAGKTEKKNYNNIWPVPPNIGSDELTLQTVIEGNLSILCE 2220  
QY 241 AT-----GS-Y-D-----G-----F-----A- 249  
Db 2221 SSGIPPPNLIWKKGSPVLTDMSGRVRLISGGRQLOISIAEKSDAALYSCVASNVAGTAK 2280  
QY 250 -----R-----R-----I-----WTKDG----- 256  
Db 2281 KEYNLQVYIRPTITNSGSHPTBIIVTRGKSISLECEVOQIPPTVTWMDKDHPLIKAGV 2340  
QY 257 -----NL-AS-----T-----L-----GOH----- 265  
Db 2341 EILDEGHILQKNIIHVSDTRGVYCVAVNVAGTMDKKYDLSVHAPPSIIGNHRSPIENVV 2400  
QY 266 -K-----G-PI-----FALK-W----- 274  
Db 2401 ERKNSVLTCEASGIELPSTTF--KDGWPSVLSNSVRILSGGRMLRMQTTMEDAQYTC 2458  
QY 275 -----N-----K-K-----GN----- 279  
Db 2459 VVRNAGEBERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPBITWHKDG 2518  
QY 280 -----F-I-----L-----S-----AGV- 286

Db 2519 QPLQDEAHIIISGRFLQINVOVPHTRGYTCCLASSPAGHSRSPSLNVFVSPTIAGVG 2578  
QY 287 -D---K-T---T-I---I-W----- 293  
Db 2579 SDGNPEDVTILNSPTSLVCBAYSYPATITWFKDGTPLSNRNIRILPGORTLIQILNAQ 2638  
QY 294 -D---A---H---TG-----E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVYIPPIINKGDLWGPGSLPKPEVKIKVNNLTLECE 2698  
QY 300 A-----K-----Q----- 303  
Db 2699 AYALPSASLWYKDGQPLKSDHVNIAANGHTLQI KEAQISDTGRTCVASNIAGEDEL 2758  
QY 304 F-----P-F-----H- 307  
Db 2759 FDNVIOVPPSFQKLWEIGNMLDTGRNGEAKDVIIINNPISLYCETNAAPPPTLTWYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315  
Db 2819 LTSKDKVLIILPGGRVLOIPRAKVEDAGRYTCVAVNEAGESLSQYDVRVLVPIIEGANS 2878  
QY 316 -----WQ-----SN----- 319  
Db 2879 LPBEVTVLNVKSALIECLSSGSPAPRNSWKDGQPLLEDHDKFLSNGRILQILNTQITD 2938  
QY 320 -----NT-----PAS-C-----S 326  
Db 2939 IGRYVCAENTAGS AKYFNLNVHVPSPVIGPKSENLTVVVNNPISLTCEVSGFPDLS 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKXQPIKLTNTLIVPGRTLOIPRAKVEDAGRYTCVAVNEAGESLSQYDVRVLVPIIEGANS 3058  
QY 332 --H-----V-C-----K-----L-GQ----- 338  
Db 3059 IKDHSBSLVNVREGTSVLECSNAPPPVITWYKNGRMITESTHVEILADGQMLHI 3118  
QY 339 -----D-----R-----P-IK----- 343  
Db 3119 KKAESVDTGYVCRAINVAGRDKNFNLNVHVPSPVIGPKSENLTVVVNNPISLTCEVSGFPDLS 3178  
QY 344 --T-----F-----Q-----G-----H-----T----- 349  
Db 3179 IPPTIAWLKXKRIENSLSLEVRILSGSKLQIARSQHSDSGNYTCIASNMEGKAQYY 3238  
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
Db 3239 FLSIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIQMLKDGKPIASGETERIRV 3298  
QY 361 --G---NLL-A-S-----C-S-----D-----D-----MTL-- 373  
Db 3299 SANGSTLNIYGALTSDTGKYTCVATNPAGBEDRIFNLNVHVPSPVIGPKSENLTVVVNNPISLTCEVSGFPDLS 3358  
QY 374 --K-----I-W-----SM-----KQ--D-----NCV----- 384  
Db 3359 TSINIECRXTGTPPPQINWLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVSN 3418  
QY 385 -----H-----Q-----D-----L-----Q----- 388  
Db 3419 RAGVDNKHVNLQVAPPNMDSMGTEITVLKGSSTSMACITDGTAPSMAWLRDQPLG 3478  
QY 389 -----Q-----H-N-----KE 393  
Db 3479 LDAHLTVSTHGMVQLLKAETEDSGKYTCIASNAGEVSKHFLKVLPEPHINGSBEHE 3538  
QY 394 I-----YT-----I--K-W----- 399  
Db 3539 ISVIVNNPLELTIASGIPAPKWTWMDKGRPLQTDQVOTLGGGEVLRISTAQVEDTGRY 3598  
QY 400 ---SPTG-----P---GT-----N----- 407

Db 3599 TCLASSPAGDDKXYLVRVHVHPNPIAGTDEPRDITVLNRQVTLCKSDAVPPVITWLR 3658  
QY 408 N-----P-----N-----ANUM-LAS-A---S-F----- 420  
Db 3659 NGERLOATPRVIRILSGRVLQINNADLGDNTANTCVASNIAGTKTREFILTVNVPNNIKG 3718  
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433  
Db 3719 GPSSLVILLNKSTVLECIAGVPTPRITWRKOGAVLAGNHARYSILENGFLHIQSAHVTD 3778  
QY 434 ---C-----I--H-----TL--T----- 439  
Db 3779 TGRYLCWATNAAGTDRRIDLVHVPSPSIAPGPTNTMTVIVNVQVTLTACEATGPKPSINW 3838  
QY 440 -K-H---Q-E-----P-V---Y----- 446  
Db 3839 RKNGHLLNVQDNQNSYRLLSSGSLVITSPSDDTATYECTVTNGAGDKRTVLTVDVQVPP 3898  
QY 447 SVA-----F-S-----P-D----- 453  
Db 3899 SIADEPTDFLVTKHAPAVITCTASGVPPFSIHWTKNGIRLLPRGDGYRILSSGAIEILAT 3958  
QY 454 ---GRY-----L---ASG--S- 461  
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLHVHBPVQPOPSSELHVILNPNPILLPCEATGTPSP 4018  
QY 462 F-----D---KCV----- 466  
Db 4019 FITWQKEGINVNTSGRHNHVLPSGGLQISRAVREDAGTYMCVAQNPAQTALGKIKLVQV 4078  
QY 467 ---H-----I-W-----NT-Q--- 472  
Db 4079 PPVISPHLKEYVIADVNDKPIITLSCADGLPPPDIITWHDGRAIVESTRQVRLSSGSLQIAF 4138  
QY 473 V-----C-----L-----HY----- 477  
Db 4139 VQPGDAGHYTCMAANVAGSSSTSTKLVHVPPRIRSTEGHYTVNENSQAILPCVADGPT 4198  
QY 478 ---L-N-----GO---VL---L---N----- 485  
Db 4199 PAINKKONVLLANLLGKYTAEPYSELILENVVLEDSGFYTCVANNAAAGDTHVSLTVH 4258  
QY 486 ---L-G--R-SIC-----L-YT-----LP-H---H--LV--- 500  
Db 4259 VLPTTPELPDGLSLNKGEQIRLS-CXATGIPLPKLTWTNNNIIPAHFDVSUNGHSSELVIE 4317  
QY 501 ---V-----I-PL-----V--- 505  
Db 4318 RVSKEDSGTYVCTAENSUGFVKATGFVYVKEPPVKGDPYPSNWIPLGNAILLNCEVKGD 4377  
QY 506 ---AL-----IE-----L---VL-----K 514  
Db 4378 PPTIQNRKGVDDIEISHRIQLNGSLAIYGTVNE 4413  
RESULT 15  
ID ADJ83137 standard; protein; 5636 AA.  
XX ADJ83137;  
XX AC  
XX DT  
XX DE  
XX DE Human hemimentin protein - SEQ ID 128.  
XX KW  
KW NOVX; cytostatic; antiarteriosclerotic; antidiabetic; antiasthmatic;  
KW antiallergic; antiinflammatory; respiratory; antiarthritic;  
KW dermatologic; antibacterial; cerebroprotective; vasotropic; cardiac;  
KW haemostatic; hypotensive; hepatotropic; neuroprotective; anorectic;  
KW neoplastic; antitumor; muscular; immunosuppressive; gynaecological;  
KW neoplastic; endocrine; ophthalmological; osteopathic;  
KW antiparkinsonian; anticonvulsant; tranquiliser; analgesic; nephrotropic;  
KW antiinfertility; antilipemic; cardiomyopathy; atherosclerosis; diabetes;



cell signal processing; metabolic pathway; asthma; allergy; emphysema;  
 autoimmunity; graft-versus-host; arthritis; cancer; stroke; haemophilia;  
 obesity; Alzheimer's; pain; chromosome mapping; tissue typing; human;  
 hemocentin.

Homo sapiens.

US2003170630-A1.

11-SEP-2003.

21-DEC-2001; 2001US-00032189.

21-DEC-2000; 2000US-0257495P.

22-DEC-2000; 2000US-0258171P.

20-FEB-2001; 2001US-0269940P.

08-MAR-2001; 2001US-0274192P.

22-MAR-2001; 2001US-0277826P.

29-MAR-2001; 2001US-0279840P.

11-APR-2001; 2001US-0282981P.

13-APR-2001; 2001US-0283656P.

31-JUL-2001; 2001US-0309247P.

10-AUG-2001; 2001US-0311754P.

17-AUG-2001; 2001US-0313331P.

(ALSO/) ALSOBROOK J P.

(TCHE/) TCHERNEV V T.

(LIUX/) LIU X.

(SPVT/) SPYTEK K A.

(ZERH/) ZERHUSEN B D.

(PATT/) PATTURAJAN M.

(LEPL/) LEPLY D M.

(BURG/) BURGESS C E.

(SHIM/) SHIMKETS R A.

(GROS/) GROSSE W M.

(SZEK/) SZEKERES E S.

(VERN/) VERNET C A M.

(LILL/) LI L.

(CASM/) CASMAN S J.

(BOLD/) BOLDOG F L.

(GORM/) GORMAN L.

(GANG/) GANGOLLI E A.

(FERN/) FERNANDES E R.

(RIEG/) RIEGER D K.

(EDIN/) EDINGER S R.

(GUNT/) GUNTHER E.

(MILL/) MILLET I.

(SCIO/) SCIORE P.

(ELLE/) ELLERMAN K.

(MACD/) MACDOUGALL J R.

(SMIT/) SMITHSON G.

Alsobrook JP, Tchernev VT, Liu X, Spytek KA, Zerhusen BD;

Patturajan M, Lepley DM, Burgess CE, Shimkets RA, Grosse WM;

Szekeres ES, Vernet CAM, Li L, Casman SJ, Boldog FL, Gorman L;

Gangolli EA, Fernandes ER, Rieger DK, Edinger SR, Gunther E;

Millet I, Sciore P, Ellerman K, Macdougall JR, Smithson G;

WPI; 2003-898249/82.

New NOVX polypeptides and nucleic acid molecules, useful for diagnosing,  
 preventing or treating NOVX-associated polypeptide disorder, e.g.  
 cardiomyopathy, atherosclerosis, diabetes, cancer, Parkinson's disease or  
 asthma.

Disclosure; SEQ ID NO 128; 263pp; English.

The invention relates to a novel isolated NOVX polypeptide. The  
 polypeptide demonstrates cytostatic, antiarteriosclerotic, antidiabetic,  
 antiasthmatic, antiallergic, antiinflammatory, respiratory,  
 antiarthritic, dermatological, antibacterial, cerebroprotective,  
 vasotrophic, cardiant, haemostatic, hypotensive, hepatotropic,  
 neuroprotective, anorectic, nootropic, antitumor, muscular,

CC immunosuppressive, gynaecological, antipsoriatic, endocrine,  
 CC ophthalmological, osteopathic, antiparkinsonian, anticonvulsant,  
 CC tranquiliser, analgesic, nephrotropic, antifertility and antilipaeamic  
 CC activities. The NOVX polypeptide, nucleic acid or antibody of the  
 CC invention may be useful for treating or preventing a NOVX-associated  
 CC disorder, such as cardiomyopathy, atherosclerosis, diabetes or a disorder  
 CC related to cell signal processing and metabolic pathway modulation.  
 CC Furthermore, the NOVX polypeptides may be useful for diagnosing, treating  
 CC or preventing diseases such as asthma, allergies, emphysema, autoimmune  
 CC disease, graft-versus-host disease, arthritis, cancer, stroke,  
 CC haemophilia, obesity, Alzheimer's disease and pain. The nucleic acids may  
 CC be used as hybridisation probes, in chromosome mapping, tissue typing,  
 CC preventive medicine or pharmacogenomics. The current sequence is that of  
 CC a protein of the invention which is related to human NOVX protein.

XX SQ Sequence 5636 AA;

Query Match 71.2%; Score 2596.5; DB 7; Length 5636;

Best Local Similarity 9.7%; Pred. No. 1.1e-30;

Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

QY 1 M-S-----I-----SS---D-----EV----- 8

Db 1 MISWEVVHTVFLFALLYSLAQDASFSQSEIRAEFEFGASTLAFVFDVTCGMYDDLQVVI 60

QY 9 -----NF-LV----- 12

Db 61 EGASKILETSLKRPKPLFNALVPHDPPEIGPVTTTDPKFKQYELRELIVGGGDCPE 120

QY 13 -----Y---R---Y---LQ----- 17

Db 121 MSIGAIAKIALEISLPGSFYVFTDARSKDYRLTHEVLTIQKQSQVFLVLTGDCDDRT 180

QY 18 -----E---SG-F-----S-H-S-----A---F 26

Db 181 IGKVVYEIATSSGQVFLDKKQVNEVLKVEEAVQASKVHLLSTHLEQAVNTWRIPF 240

QY 27 -----T-----FG-----I-----K 31

Db 241 DPSLKEVTVLSGSPSMIEIRNPLGKLKKGFGHLLHNHNSAKVNVNKEPEAGMTVK 300

QY 32 ---S-H-----I-----S-Q----- 36

Db 301 TSSGRHSVRITGLSTIDFRAGFSRKPTLDFKKTVSQVQGIPTVYLLNTSGISTPARID 360

QY 37 -----S-----NI-----N-----G----- 41

Db 361 LLELSISGSSSLKTIPTVKYYPHKKPYGIWNISDFVPPNEAFFLKVTGYDKDYLFORVSS 420

QY 42 -----A-----L---VP-----P-----A 47

Db 421 VFSSTIVDPAPKVTMPKPTPGYVLPQGPICSDVSLPPTLSFVRNGVTGLVDQYLKESA 480

QY 48 -----A---L-----I-----S---IIQ-----KG----- 56

Db 481 SVSLDIKVTLSDEGFEYCIASVSSAGTGRAQTFDVSPPVQVPPNNTVTGGERAVLT 540

QY 57 -----L---Q-----Y---V---E----- 61

Db 541 CIIISAVDNLWQRNDRVRLAEPARITLANLSLELKSVPKNDAGEYHCVMSSEGGSS 600

QY 62 -A-----EVSII-----NE----- 68

Db 601 AASVFLTVQEPKVTVMKQSQFTGSEVSIIMCSATGYPKPKIANTWNDFIVGSHRYRM 660

QY 69 --DGLTF-----D-----G-----R---P-----I-- 78

Db 661 TSDGTLFIKNAAPKADAGIYGLCLASNSAGTDKQNSTLRYIEAPKLMVQSELLVALGDITV 720

QY 79 -E---S-----L-S-LI-----DA-----V----- 87

Db 721 MECKTSGIPPPQVKKFKGDLRLRPSTFLIIDPLLLGLLKLTQETODLDAGDYTCVAINAAGR 780

QY 88 -----M-----P-DV-----VO-----T----- 94  
Db 781 ATGKITLDVSGPPVFIQEPADVSMEIGSNVTLPCYQGYPEPTIKWRRLDNMPFSRPPS 840  
QY 95 -----R----- 95  
Db 841 VSSISQLRTGALFILNLWASDKGTIYCEAENQFGIQSETTVTVTGLVAPLIGISPSVAN 900  
QY 96 -----QQ-----A-----Y-----R-----D--KL-- 103  
Db 901 VIEGQQLTLPCTLLAGNPPIERRWIKNSAMLLQNPYITVRS DGLSHIERVQLQDGGEYTC 960  
QY 104 -----A-----Q--QQ-----AA----- 109  
Db 961 VASVAGTNNKTTTSVVHVLPFTQHGOQILSTIEGIPVTLPCASGNPKPSVWSKKGEL 1020  
QY 110 -----AAA--A--A-----A--AA--A----- 118  
Db 1021 ISTSAKFSAGADGLYWSPEGESEGYVCTATNTAGYAKRVQLTVYVRPVFGDLRG 1080  
QY 119 -SQ-----O----- 123  
Db 1081 LSQDKPVEISVLAGEVTLPCFVKSLPPPIITWAKETQLISPPSRHTPLPSCSMKITET 1140  
QY 124 -----A-----K-----N----- 126  
Db 1141 RTSDSGMYLCVATNIAGNVTAQKLVNHPKIQRGPKHLKVQVGQVDPICNAQGTPLP 1200  
QY 127 -----GE--N-----T--A--N--G----- 133  
Db 1201 VITWSKGGSTMLVDGHHVSNPDGTLISDQATPSDAGIYCVATNIAGTDETEITLHVQE 1260  
QY 134 -----E--N--N-----GA-----H----- 139  
Db 1261 PPTVEDLEPPYNTTFQERVANQRIEPPCPAKGTPKPTIKWLHNGRELGTREPGISILEEG 1320  
QY 140 T-----IA-----N-----N--HT-----DM----- 148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKQEQVNSVLLNQLTNL 1380  
QY 149 -----M-----E----- 150  
Db 1381 FCEVEGTPSPIIMYKDNQVQTESSTIQTVNGKILKLFATPEDAGRYSCAKAINIAGTS 1440  
QY 151 -----VD-----G-----DV--E-----I-----P-----PN 160  
Db 1441 QKYFNIDVLPPTIIGTFPKESVVLNRDVALEQCQVKGTPFPDIHWFQDGLFLGDPN 1500  
QY 161 -----K--A-----V-----V-----VL 165  
Db 1501 VELLDRGQVLHLKNARNDKGRYQCTVSNAAQQAQDKLTIYIIPPSIKGNVTTDISVL 1560  
QY 166 -----RG----- 168  
Db 1561 INSLIKLECKTRGLPMPAITWYKDGQPIIMSSQALYIDKQYLHIIPRAQVSDSATYTCV 1620  
QY 169 -----E--S-----E-----VFI-----C--A-----W----- 177  
Db 1621 ANVAGTAESFHVVDVYVPPMIEGNLATPLNKQVVIASHLTLECNAGNPSPIITLWKDGV 1680  
QY 178 -----N-----P--V--SD-- 182  
Db 1681 PVKANDNFRIEAGGKKLEIMSAQIDRGQYICVATS VAGEKEIKYEDVDVLVPPAIEGGDE 1740  
QY 183 -----LL-----ASGS-----G-----D-----S-- 191  
Db 1741 TSYFIWVNNLLELDCHVTGSPPTIMWLKQGLIDERDGFKILLNKRKLVIQAQVSN 1800  
QY 192 -----TA-----R-----I--W 196  
Db 1801 GLYRCMAANTAGDHKKEFEVTVHVPPTIKSSGSLSERVVVKYKPVALQCIANGIPNPSITW 1860  
QY 197 -----NL----- 199

Db 1861 LKDDQPWTAQGNLKITQSSGRVLQIAKTLLEDAGRYTCVATNAAGETOQHQLHVHPPS 1920  
QY 200 -E-----N-----S-----T-----SGST----- 207  
Db 1921 LEDAGKMLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLLSGTSMTPLNRGQIIDIES 1980  
QY 208 -Q-----L--V-----LR-H-----C--I-- 215  
Db 1981 AQISDAGIKYCVAINSAGATELFYSLQVHVAPSISGNNMVAVVVANNPVRLECEARGIPA 2040  
QY 216 -----R-----E--G-----G--O--D-- 221  
Db 2041 PSITLWKDGPVSSPENGLOVLSGGRILALTSTQISDTGRYTCVAVNAAGEKORDILRV 2100  
QY 222 -VPSN-----KD-----VT--S--L-- 231  
Db 2101 YVPPNIMBEEQNVSVLISOAVELLQSDAIPPTLTWLDKGHPLLKPKGLSISENRSVLK 2160  
QY 232 -----D-----W--N--S-----EG--TLL-- 240  
Db 2161 IEDAQVODTGRYTCREATNVAGTEKKNYNNIHWPPNIGGSDELQTLTVIEGNLISLCE 2220  
QY 241 AT-----GS--Y--D-----G-----F-----A-- 249  
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVRLSSGGRQLQISIAEKSDAALYSCVASNVAGTAK 2280  
QY 250 -----R-----I-----WTXDG----- 256  
Db 2281 KEYNLQVIRPTITNSGSHPTETIIVTRGKSISLECEVOGIPPTVTWMDKGHPILKAKGV 2340  
QY 257 -----NL--AS-----T--L--L-----GOH----- 265  
Db 2341 EILDEGHILQKNIHVSDTGRYVCVAVNAVGMTDKKYDLSVHAPPSIIGNHRSFENISVV 2400  
QY 266 -K-----G--PI--FALK--W----- 274  
Db 2401 EKMSVSLTCEASGIPLPSTWTF--KDGWPVLSNSVRILSGGRMLRMQTTMEDAGQYTC 2458  
QY 275 -----N-----K--K-----GN----- 279  
Db 2459 VVRNAAGEERKIFGLSVLPVPHIVGENTLEDVKVEKQSVTLTCEVTGNPVPBITWHKDG 2518  
QY 280 -----F--I-----L--S--AGV-- 286  
Db 2519 OPLQDEAHIISSGRFLQITNVQVPHTCRYTCLASSPAGHKRSFSLNVFVSPTIAGVG 2578  
QY 287 -D--K--T-----T--I-----I--W----- 293  
Db 2579 SDGNPEDVTILNSPTSLCEAVSYPPATITWPKDGTPLSNRNRILPGGRTLQILNAQ 2638  
QY 294 -D-----A-----H-----TG-----E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKYVYIPPIINKGLWPGPLSPKEVKIKVNNLTLECE 2698  
QY 300 A-----K-----Q--O----- 303  
Db 2699 AYAI PSASISLWYKDGQPLASDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758  
QY 304 F-----P--P-----LDV-----D 315  
Db 2759 FQVNIQVPPSFQKWEIGNMLDTRNGEAKVDIINNPIISLYCETNAAPPPTLTWYKDGHP 2818  
QY 308 --SA-----P-----A----- 319  
Db 2819 LYSKDKVILPGGRVLQIPRAKVEDAGRYTCVAVNEAGDSLQYDVRVLVPPPIEGANS 2878  
QY 316 -----WQ-----SN----- 319  
Db 2879 LPEEVTVLVKNKALIECLSSGSPAPNSWQKQDQPLEDDHKKFLSNRILQILNTQITD 2938  
QY 320 -----NT-----FAS--C-----S 326

Db 2939 IGRYVCAENTAGSAXKYFNLNVHVPSPVIGPKSENLTVVVNNFISLTCEVSGFPDPDLS 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKXQPIKLTWTLVPGGRTLOIRAKVSDGGEYCTAINXAGSKKKFSITVVVPPS 3058  
QY 332 H-----V-----C-----K-----L-GQ----- 338  
Db 3059 IKDHSESLSVNVREGTSVLECESNAVPPVITWYKNGRMITESTHVEILADGQMLHI 3118  
QY 339 -----D-----R-----P-IK----- 343  
Db 3119 KKAESDGTGOYVCRAINVAGRDDKNFHLNVVYPPSIEGPEREVIVTISNPVTLTCDATG 3178  
QY 344 -----T-----F-----Q-----G-----H-----T----- 349  
Db 3179 IPPPTTAMLKNYKRIENSLSVRILSGGSKLQIARSQSDSGNYCTIASNMEGKAQYY 3238  
QY 350 -----N-E-V-NA-----IKW-D-P-----T----- 360  
Db 3239 FLSIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIQWLDGKPIASGETERIRV 3298  
QY 361 -----G-----NLL-A-S-----C-S-----D-----D-----MTL-- 373  
Db 3299 SANGSTLNIYGALTSYGKYTCVATNPAGEEDRIFNLNVVYPTIRGNKDEAKLMTYVD 3358  
QY 374 -----K-----I-W-----SM-----KQ-D-D-----NCV-- 384  
Db 3359 TSINIECRXTGTPPPQINWMLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418  
QY 385 -----H-----Q-----D-----L-----Q----- 388  
Db 3419 RAGVDNKHYNLQVFAFPNDMSMGTEBITVLKGSSTSMACITDGTAPSMAWLRDGOPLG 3478  
QY 389 -----Q-----H-N----- 393  
Db 3479 LDAHLTVSTHGMVQLLKAETEDSGKYTCIASNEAGEVSKHFLKVLPEPPHNGSEHEE 3538  
QY 394 I-----YT-----I-K-W-----N----- 399  
Db 3539 ISVIVNPLELTCTIASGIPAPKMTWMDGRPLPOTQOVTLGGGEVLRISTAQVEDTRY 3598  
QY 400 -----SPTG-----P-----GT-----N----- 407  
Db 3599 TCLASSPAGDDDKKEYLVRVHVHPNIAGTDEPRDITVLRNQVLTCKSDAVPPPVITWLR 3658  
QY 408 N-----P-----N-----ANLM-LAS-A-----S-----F----- 420  
Db 3659 NGERLOATPRVIRILSGGRYLQINNADLGDYNTATYCVASNIAGKTTREFILTVNVPPNIKG 3718  
QY 421 -----DSTV-----RL-W-D-V-----DR-----G-I----- 433  
Db 3719 GPQSLVILLNKSTVLECIAGVPTPRITWRKOGAVLAGNHARYSILENGFLHIQSAHVTD 3778  
QY 434 -----C-----I-----H-----TL-----T----- 439  
Db 3779 TGRYLCMATNAAGTDRRIDLOVHVPPSIAPGPTNMTVIVNVQTLTACEATGIPKPSINW 3838  
QY 440 -K-H-----Q-E-----P-V-----Y----- 446  
Db 3839 RKNHLLNVNDONQSYRLSSGSLVLIISPSVDDTATYECTVTNGAGDKRKTVDLTQVPP 3898  
QY 447 SVA-----F-S-----P-D----- 453  
Db 3899 STADEPTFLVTKHAPAVITCTASGVPPFPIHWTNKGIRLLPRGDGYRILSSGAIEILAT 3958  
QY 454 -----GRY-----L-----ASG--S- 461  
Db 3959 QLNHAGRYTCVARNAAGSAHRVTLHVHPPVIOQPQSELHVLNNPILLPCEATGTPSP 4018  
QY 462 F-----D-----KCV----- 466  
Db 4019 FITWQKGINVNTSGRHAFLPSGGIQLISRAVREDAGTYMCVAQNPAQTALGKILNVQV 4078

QY 467 -----H-----I-W-----NT-Q--- 472  
Db 4079 PPVISPHLKEYVIAVDKPIITLSCADGLPPDITMHKDGRAIVESIRQVRLSSGSLQIAF 4138  
QY 473 V-----C-----L-----HY----- 477  
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTVHVPPRIRIRSTEGHYTVNENSQAILPCVADGIPT 4198  
QY 478 -----L-N-----GO-----VL-----L-----N----- 485  
Db 4199 PAINWKKONVLLANLLGKYTABPYGELIILENVVLEDSGFYTCVANNAAAGEDTHVSLTVH 4258  
QY 486 -----L-G-----R-SIC-----L-YT-----LP-H-----H-LV-- 500  
Db 4259 VLPTFTLPGDVSLNKGEOQLRLS-CKATGILPLPKLTWTFNNNIIPAHFDSVNGHSELVIE 4317  
QY 501 -----V-----I-PL-----V----- 505  
Db 4318 RVSKEDSGTYVCTAENSVGFKATGFVYVKEPVFKGDYPSNWIPLGGNAILNCEYKGD 4377  
QY 506 -----AL-----IE-----L-----L-VL-----K 514  
Db 4378 PTFTIOWNRKGVDDIEISHRIRQLNGSLAIYGTVNE 4413

Search completed: January 3, 2005, 15:19:58  
Job time : 116.333 secs

This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:10:17 ; Search time 23.6667 Seconds  
(without alignments)  
1440.315 Million cell updates/sec

Title: US-09-987-701-4

Perfect score: 3649

Sequence: 1 MSISDEVNFLVRYLQESG.....LPHLVIVFLVALIELVLK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents AA:\*
- 1: /cgn2\_6/ptodata/1/iaa/5A COMB.pcp.\*
  - 2: /cgn2\_6/ptodata/1/iaa/5B COMB.pcp.\*
  - 3: /cgn2\_6/ptodata/1/iaa/6A COMB.pcp.\*
  - 4: /cgn2\_6/ptodata/1/iaa/6B COMB.pcp.\*
  - 5: /cgn2\_6/ptodata/1/iaa/PTUS COMB.pcp.\*
  - 6: /cgn2\_6/ptodata/1/iaa/backfile1.pcp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2571.1	70.5	3461	3	US-09-334-220-2
2	2562.6	70.2	4544	1	US-08-469-486-52
3	2562.6	70.2	4544	2	US-08-469-658-52
4	2558.9	70.1	4654	3	US-08-476-515A-84
5	2558.8	70.1	4655	3	US-08-652-877-84
6	2558.8	70.1	4655	3	US-08-652-877-88
7	2558.8	70.1	4655	3	US-08-652-877-90
8	2557.8	70.1	4655	3	US-08-652-877-86
9	2554	70.0	3460	3	US-09-334-220-1
10	2552.2	69.9	5405	3	US-08-718-388-9
11	2533.4	69.4	15281	2	US-08-471-119A-2
12	2527.7	69.3	4861	4	US-09-919-497-70
13	2527.4	69.3	4303	2	US-08-460-751-2
14	2527.4	69.3	4303	4	US-09-479-467A-2
15	2526.5	69.2	4302	3	US-09-052-469-8
16	2526.5	69.2	4302	4	US-08-422-582-8
17	2526.5	69.2	4302	4	US-09-052-262-8
18	2526.5	69.2	4339	3	US-09-052-469-6
19	2526.5	69.2	4339	4	US-08-422-582-6
20	2526.5	69.2	4339	4	US-09-052-262-6
21	2523.5	68.8	3571	4	US-09-911-842A-2
22	2511.4	68.8	4302	3	US-08-658-136-5
23	2477.3	67.9	3594	2	US-09-911-842A-4
24	2463.6	67.5	3778	2	US-08-222-617A-2
25	2462.1	67.5	4866	4	US-09-424-783-2
26	2460.1	67.4	3727	2	US-08-222-617A-27
27	2459.4	67.4	3959	2	US-08-970-269A-30

28	2459.4	67.4	3959	3	US-09-407-562-30	Sequence 30, Appl
29	2459.1	67.4	3666	2	US-08-222-617A-12	Sequence 12, Appl
30	2456	67.3	4928	3	US-09-036-987A-5	Sequence 5, Appl
31	2456	67.3	4928	3	US-09-370-700-5	Sequence 5, Appl
32	2456	67.3	4928	4	US-09-603-207-5	Sequence 5, Appl
33	2451.4	67.2	7257	3	US-09-335-409-5	Sequence 5, Appl
34	2451.4	67.2	7257	3	US-09-568-102-5	Sequence 5, Appl
35	2451.4	67.2	7257	3	US-09-567-969-5	Sequence 5, Appl
36	2451.4	67.2	7257	3	US-09-568-480-5	Sequence 5, Appl
37	2451.4	67.2	7257	3	US-09-568-486-5	Sequence 5, Appl
38	2451.4	67.2	7257	3	US-09-568-472-5	Sequence 5, Appl
39	2451.4	67.2	7257	3	US-09-567-899-5	Sequence 5, Appl
40	2434.6	66.7	4536	4	US-09-079-030-1	Sequence 1, Appl
41	2434.6	66.7	4563	4	US-09-538-092-842	Sequence 842, App
42	2433.9	66.7	5032	4	US-09-538-092-979	Sequence 979, App
43	2433.6	66.7	4563	4	US-09-108-006C-1	Sequence 1, Appl
44	2433.5	66.7	3898	3	US-08-750-717-2	Sequence 2, Appl
45	2432.6	66.7	4536	4	US-09-180-422B-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1  
US-09-334-220-2  
; Sequence 2, Application US/09334220  
; Patent No. 6323177  
; GENERAL INFORMATION:  
; APPLICANT: St. Jude's Children's Research Hospital  
; APPLICANT: Curran, Thomas  
; APPLICANT: D'Arcangelo, Gabriella  
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
; TITLE OF INVENTION: THERAPIES  
; FILE REFERENCE: 2427/0F704  
; CURRENT APPLICATION NUMBER: US/09/334,220  
; CURRENT FILING DATE: 1999-06-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 3461  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-334-220-2

Query Match	70.5%	Score	2571.1	DB	3	Length	3461
Best Local Similarity	11.5%	Pred. No.	66-33				
Matches	396	Conservative	94	Mismatches	16	Indels	2939
		Gaps	345				
QY	1	M	-----SI-----S-----S-----S-----D-----EV-----8				
Db	1	M	-----SI-----S-----S-----S-----D-----EV-----8				
QY	9	N	-----N-----F-----LV-----YRY-----LQ-----ES-----GF-----S-H 23				
Db	61	H	-----N-----F-----LV-----YRY-----LQ-----ES-----GF-----S-H 23				
QY	24	S	-----S-----A-----F-----T-----T-----F-----G-----I-----K-----31				
Db	120	Q	-----S-----A-----F-----T-----T-----F-----G-----I-----K-----31				
QY	32	S	-----S-----A-----F-----T-----T-----F-----G-----I-----K-----31				
Db	180	E	-----S-----A-----F-----T-----T-----F-----G-----I-----K-----31				
QY	44	V	-----P-----P-----A-----AL-----I-----SI-----IQ-----54				
Db	240	A	-----P-----P-----A-----AL-----I-----SI-----IQ-----54				
QY	55	K	-----KG-----LQY-----V-----EA-----E-----VSIN-----67				
Db	300	K	-----KG-----LQY-----V-----EA-----E-----VSIN-----67				

QY 68 ---ED---GT-LF---DG---R---76  
Db 359 VVLEDNLDPVDTGNWLFPPGATVKHSCQSDGNSIYFHGNEGSEFNATTRDVLSTEDIQ 418  
QY 77 ---P---I---ES-LSL1---DA---86  
Db 419 EQWSEFESQPTGWDILGAVVGADCGTVESGLSVFLKODGERKLCPTPYMDTTGYGNLRFY 478  
QY 87 -VM---P---D---VW---QT---94  
Db 479 FVMGGICDPCGVSHENDIILYAKIEGRKEHIALDTLTYSYKVPSPVSVINPELQTPATK 538  
QY 95 ---RO---QAY-R---D---K-LAQ---Q---106  
Db 539 FCLQXSHQGYNRVWAVDFHVLVPLPSTMSHMIQPSINLGCCTHQPNGSVSLEFSTNH 598  
QY 107 -QA---A---AAA---AA---114  
Db 599 GRWSLLHTECLPEICAGPHLPSTVYSSENYSGMNRITIPLPNAALTRDTRIRWRQTGP 658  
QY 115 ---A---A---AASQ---Q-GS 123  
Db 659 ILGNMWAIDNVYIGPCLKFCSGRGQCTRHGCKDFGSGFACEMASQTFPMFISESFGS 718  
QY 124 AK---N---G-E---N---129  
Db 719 ARLSSTHNFYSIRGAEVSGCGVLASGKALVFNKGRQRLITSPDSSQSRFLQFTLRIG 778  
QY 130 ---T-A---NG---E-E---N---G-137  
Db 779 SKSVLSTCRAPDQGEVLLHYSDNGITWKLLEHYSYVNYHPRISVLPDARQFGI 838  
QY 138 ---A-H---T---I---AN---N---H-145  
Db 839 QFRWQPYHSSQGEDVWAIDEIVMTSVLFNSISLDETNLVEVTQSLGFLGNVQPYCGHD 898  
QY 146 ---T-D---M---NEVDG---D-V-E---156  
Db 899 WTLCTGDSKLASMRVETQSMQI-GASYMIQFSLVMGCGQKYTHMDNQVKLEYSANH 957  
QY 157 ---I---P-PN-K---A---VVL---R---166  
Db 958 GLTWHLVQBECLFSMFCQFTSASIVHASEFTQWRRTVVLPQKTWSGATRFMSQSY 1017  
QY 167 ---CH---E-SE---V-P---173  
Db 1018 TAQDEWALDNIYIGQCPNMCSGHGCDHGVCRCDOGYQTECHPEAALPSTIMSDFENP 1077  
QY 174 ---I---C---A---W---177  
Db 1078 SSWESDQEVIGEVVKPEGCGVSVSGSSLYFSKAGKROLVSWDLDTSWDFVQFYQI 1137  
QY 178 ---N-P---V---SD---L-L-ASG---187  
Db 1138 GGESAACNKPDSREEGILLQYSNNGGIQWHLAEMVFSDFSKPRFVYLELPAAGKTPCTR 1197  
QY 188 ---SG---D---S---T---A---R-IW-196  
Db 1198 FRWKPVSGEDYDQWAVDDIILSEKQKQVIVPNPTLPQNFYKPAFDYPMNQMSWL 1257  
QY 197 ---N-LSEN-S---T---S---G---ST 207  
Db 1258 MLANEGMAKNDSFCAITPSANVFGKSDGDRFAVTRDLTKPGVYLQFKLNGICTSQSST 1317  
QY 208 ---Q---L---V---210  
Db 1318 APVLQYSHDAGMSWFLKEGCFPASAAGCEGNSRELSPTVYVYTGDPBEWTRITIAIP 1377  
QY 211 ---L---R---H-CI---215  
Db 1378 RSLASSKTRFRWIQESSQKNVPPFGLDGYVIBEPCPSYCSGHGDCISGVCFDLGYTAA 1437  
QY 216 ---R-E---GGQ---D---V 222

Db 1438 QGTCVSTNTHSEMDFRFSGLKPLWYKITGQVGTGCTLNDGRSLYFNLGKREARTV 1497  
QY 223 P---SN---K-D---227  
Db 1498 PLDTRNISLVQFYIQIGSKTSGITYITPRARYEGLVVOVYSNDNGILMHLRLDFMSFLE 1557  
QY 228 ---VTSLD---W---N-SE---236  
Db 1558 POIISIDLPREAKTPATAPRWQPOGHKHSQWALGDVLIGVNDSSQTFQDKLGSIDL 1617  
QY 237 ---G---T-L---LA-T---G---243  
Db 1618 QANWYRIQGGQVDIDCLSMDTALITFTENIGNPRYAETWDFHVSSESLQWMMNGCKP 1677  
QY 244 S---Y---D---G---FA-RI---251  
Db 1678 SGAHGILQYSLNMGKDWLVTEECVPPTIGCVHYTESSTYTTSERFQNRRTVYLPLAT 1737  
QY 252 ---W---T---253  
Db 1738 NSPRTFRWIOQNYTVGADSWAIDNVILASGCPWMCGRGICDSGRVCDCRGGFGFCVP 1797  
QY 254 ---KD---GNL---A---S-T-L---G---263  
Db 1798 VVPLSILKDDFNGNLHPLDLPPEVYGAERGNLNGETIKSGTCLIFKGEGLRMLISRDLDC 1857  
QY 264 ---Q---HK---G---P---I-F---270  
Db 1858 TMTYVQFSRLRTAKTGPERSHILLOQSVSGVTWHLMDFFYFPQTTSILFINVPLPYG 1917  
QY 271 ---A---LKW---N-KK---GN---F---280  
Db 1918 AQTNAFRFL-WQPNNGKKEIWIIDDFIIGNNLPVLLDITDFGPRDNWFFYPG 1976  
QY 281 ---I-L---SA---G-V---D---KTT11---292  
Db 1977 GNIGLYCPYSSKGAPEEDSAMVFSVNEVGEHSITTRDLSVNENTIIOFEINVGCSTDSS 2036  
QY 293 ---W---DA-H---T-G---EA---300  
Db 2037 ADPVRLEFRDFGATWHLLPLCYHSSSLVSLCSTEHPSTYYAGTTQGMRRVVHFG 2096  
QY 301 K---Q-O-F---P---F-H---307  
Db 2097 KLHLCGSVFRVYQGYFAGSQPVTAIDNVYIGPOCEMCGHSGCINGTKICDPYS 2156  
QY 308 ---SA---PA---311  
Db 2157 GPTCKISTKNPDKDDFSGQLESDFLLMSGGKPSRKCGILSSGNNLPFNEGLRLMT 2216  
QY 312 ---LD---V-D---315  
Db 2217 RDLDLSHARFVQFMRLGCGKGVDPFRSQPVLLQYSLNGLSWLLQEFLLFSNSNVGRY 2276  
QY 316 ---WO---S---N-NT---321  
Db 2277 IALEMPLKARSGSTRLRWQPSENGHPYFPWIDQILIGNISGNTVLEDDFSTLDSRW 2336  
QY 322 ---F-ASCS-TDMC---I 331  
Db 2337 LLHPGQTKMPVCGSTGDALVFIEKASTRYVVTDDIADVNEBSFLQIDFAASCSVTDSYAI 2396  
QY 332 ---H---V-C---334  
Db 2397 ELEYSVDLGLSWHPLVRDCLPTNVECSRYHLQIRILVSDTFNKNWTRITLPLPSYTRSOATR 2456  
QY 335 ---K---L---G-Q---D-R---P---IK 343  
Db 2457 FRWHPAPFDKQOTWADNVYIGDGLDMCSGHGRCVQSCVDEQWGGLYCDEPEPISLP 2516  
QY 344 T---F---Q---G---H---TN---350

2517 TOLKDNFRAPSNQNLVTVGGKSLTVCGAVASGLALHFGGCSRLLVTVLNLNTNAETI 2576  
351 -----E-VN-AIKW-----D-----P-----359  
2577 QFYFMYGCLITPSNRNQVLLVSVNGGITWNLMEIFDQYKSGFVNILLPPDAKEIA 2636  
360 T-----G-----N-L-----L-----364  
2637 TRPRWQPRHGLDQNDWADNLVIGSADQRTVMDLTFSSAPVQHERSPADAGPVGR1 2696  
365 A-----S-----CS-----D-D-M-----T-L-----373  
2697 AFEMFLEKTSVNNENLWFLHDDCTVERFCDSPDGMVLCGSHDGRVAVTHDLTPFENWIM 2756  
374 -----KI-----375  
2757 QPKISVGCCKVPEKIAQNIHVQSTDFGVSWSLVPOCLPADPKCSGSVSQSPVFPFTEG 2816  
376 W-----SM-----K-----Q-----DN-----CV-----H-D-L-QQ-----389  
2817 WKRTIPLPESLTGNPVRFRFYQKYSQVQWADIDNFGVLPGLDNCGHGDCLEKQICDP 2876  
390 -----H-N-K-----EI-----Y-----T-----396  
2877 GYSGPNCYLTHSLKTLKRFDEEIKPDLWMSLEGSTCTCGVLAENTALYFGGSTVR 2936  
397 -I-----K-----W-----S-P-----T-G-----403  
2937 QAITQDLRGAKPLQVWGRIGSENWNTSCHRVCCKEGLVLDFTDGGITWLLHEMDF 2996  
404 -----P-G-TNN-----P-----N-----A-NLML--A-----416  
2997 QKVISVRHDYILLPEGALITWTLRWQPPVINGLVSVGVERAQWALDNLILGAEINP 3056  
417 SA--SFD--ST-----VR-----LM-----D-----V-DR-----431  
3057 SOLVDTFDDGSGSHEENWSPYFNAVATAGFCGNPSFHLWPNKKDKDTHNLSRELLIQ 3116  
432 -G-----I-----C-IHT-----LTK-----H-----441  
3117 PGYMMQFKIVGGEATSCGDLHSLVMEITYTKDARSQNLVQTOCLPSSNSISCSPPQPH 3176  
442 QE-PVY-----S-----VAFSPD-----G-----R-----YL-A-----458  
3177 -EATYNAVNSSWKRTIQL-FDHVSSSATQFRWIKQKEETEKQSWAIDHVVIGACPK 3234  
459 --SG-----SF-DKC-V-H-----I-----WNT-Q-V-----473  
3235 LCSGHGYCTTGAVCICDESPQGDGCSVFSHELPSYIKDNFESARVTEANWETIQGGVIGS 3294  
474 -C-L-----H-----YLANG-QV-----L-LN-----L-G--R-SIC--LYTL 495  
3295 GCGQLAPYAHGDSLYFGCGQIROAATKPLDLTRASKIMFVLOTGSPAQTDG-CNSDL-SG 3352  
496 PH-----HLV-----V-----IPL-----V-----505  
3353 PHTVDKAVILQYVNNNGITWHTVIAHQPKDFTQARQSVYNVPLEARMKGVLLRWQPREN 3412  
506 -----AL--IE-LLV-----LK 514  
3413 GTGHDQWALDHVEWLVSTRKQNYM 3437

## RESULT 2

US-08-469-486-52  
; Sequence 52, Application US/08469486  
; Patent No. 5739281  
; GENERAL INFORMATION:  
; APPLICANT: Thoegeersen, Hans Christian  
; APPLICANT: Holtet, Thor Las  
; APPLICANT: Etzerodt, Michael  
; TITLE OF INVENTION: Improved method for the refolding of  
; TITLE OF INVENTION: proteins

NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson  
STREET: 225 Franklin Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02110-2804  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version  
SOFTWARE: #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/469,486  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/192,060  
FILING DATE: February 4, 1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Paul T. Clark  
REGISTRATION NUMBER: 30,162  
REFERENCE/DOCKET NUMBER: 06363/002001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617 542 5070  
TELEFAX: 617 542 8906  
TELEX: 200154  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4544 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-469-486-52

Query Match 70.2%; Score 2562.6; DB 1; Length 4544;

Best Local Similarity 10.4%; Pred. No 3 9e-32;

Matches 414; Conservative 79; Mismatches 16; Indels 3474; Gaps 366;

QY 1 MS-----I-----SSDE-----V-----N-F-L-----11  
DB 88 MSRLNGVQDCMDGSDGPHCHRELQNCGRSGHCHVPTLDGPTCVCSNPFOLQADGKT 147  
QY 12 -----VY-----R-----Y-LQ-----E-----18  
DB 148 CKDFDECSYVGTCSQLCTNTDGSFICGCVGYLLQPDNRSCKAKNEPVDPRPVLIIANSQ 207  
QY 19 -----SG-----FS-----H--SA-----F 26  
DB 208 NILATVLSAQVSTTPTSTROTAMDVSANETVCWVHVGDSAAQTOLKCARMPGLKGF 267  
QY 27 -----TEGIK-S-HISQ-----S-N-----I-----NG-----41  
DB 268 VDEHTNISLSLHHVQMAIDMLTGNFYFVDDIDDRIFVCRNRNGDTCVTLDDLELYNPKG 327  
QY 42 -AL-----VP-----P-----A-A-48  
DB 328 IALDPAMGVFFTDYQGIKPKVERCDMDGQNRKLVDSKIVFPFHGITLVLVSLVYWADAY 387  
QY 49 --LJ-----SIIQK-L-Q-Y--V-E-----A-EVS-I-----N 67  
DB 388 LDYIEVDVYEGKGRQTIIO-GILIEHLYGLTVFENLYATNSDANAAQKTSIVNRFN 446  
QY 68 --E-----D-GTL-----R-----72  
DB 447 STIEYQVTVTRVCKGALHIYHQRQPRVSHACENDQYKGGCGDCLLANSKHARTCRC 506  
QY 73 ---F---DG-----R-----PIE-----SL---81  
DB 507 RSGFSLGSDGSKCKPEHELFLVYKGRPGCIIRGMDMGAKVDPDEHMIPIENLMPRALDF 566

QY 82 -----S-LI-----D-----AV--M-----PD-- 90  
Db 567 HAETGYFADTTSYLIGROKIDGTERETILKOGIHNVGVAVDMGDNLYWTDGPKKT 626  
QY 91 V-V-----QTR-----QO----- 97  
Db 627 ISVARLEKAAQTRKTLIEGKMTWHPRAIVDPNLNGMYMTDWEEDPKDSRRGRLEAMWDG 686  
QY 98 -----A-Y-R-----D-K----- 102  
Db 687 SHRIDFTSVTLWPNGLSLDIPAGRLYWDAFYDRIETILLNGTDRKIVVEGPELNHAF 746  
QY 103 -----L-----AQOAA----- 109  
Db 747 GLCHGNYLFWTEYRSGSVYRLERGVGAPPTVTLRSEPPPIFEIRMYDAQOQOVGTNK 806  
QY 110 -----AA-----A-A-----A-AASQ- 120  
Db 807 CRVNNCGSSILCLATPGSRQCAEAEDQVLDADGVTCLANPSYVPPQCPQGFACANSRC 866  
QY 121 Q-----G-----S-----K-----N-G-----E 128  
Db 867 IOERWKCDGNDCLDNDSEAPALCHQHTCPDRFKCENNRCPNRLWLCGDNDNCGNSEDE 926  
QY 129 N-T--A-----N-----G-----E-----E--N 136  
Db 927 SNATCSARTCPPNQFSCASGRCIPISWTCBLDDDCGRSDSASCAYPCTCPPLTQFTCNN 986  
QY 137 G-----A--H-----T-----I----- 141  
Db 987 GRCININWCDNDNDGNDSEAGCSHSCSTQPKNSGRCPHEWTCDCGNDNDGYSDE 1046  
QY 142 --AN--N-----HTD-----M-----M--E-----V-D- 152  
Db 1047 THANCTNQATRPGGCHTDFQCRLDGLCPLRWCRDGTDCMDSDSEKSCGVTHVCDP 1106  
QY 153 --G--D-----V-----E-I--PP-----NKAV-----V- 164  
Db 1107 SVKFGCKDSARCISKAWDCGNDNDCEDNDEBNCESLACRPPSHPCANNTSVCLPPDKLC 1166  
QY 165 -----L--R-----G--H-----E--S-----E-----V-- 172  
Db 1167 DGNDDCGDSGDEGLCDQCSLNGGCSHNSVAPGEGIVCSPLGMBELGPDNHTCQIQSY 1226  
QY 173 -----F-I-CA--W-----N----- 178  
Db 1227 CAKHLKCSQKCDQNKFSKVCYEGWLEPDGSCSLDPFKPFIIFSNRHEIRRIDLHK 1286  
QY 179 -----P-----V-----S-----D-----LL--A--S----- 186  
Db 1287 GDYSVLVPGURNIATLDFHLSQSALYWTVDVVEDKIYRGKLLDNGALTSFEWVIQYGLATP 1346  
QY 187 -----G-----S-----GD-----S-----T-- 192  
Db 1347 EGLAVDWIAGNIYVESNLQIEVAKLDGTLRTLLAGDIEHPRAIALDPRDGLFWTDW 1406  
QY 193 A--RI-----W-N----- 197  
Db 1407 DASLPRIEAAASMGAGRRTVHRETGGGWPNGLTVDYLEKRILMIDARSARYDGS 1466  
QY 198 -----LS-----E-----NST-----SG-----S--TQ--L 209  
Db 1467 GHMEVLRGHFLSHPPFAVTLYGGEVYWTDRWTN--TLAKANKWTGHNVTVVQRTNTOFPDL 1525  
QY 210 -V-----L-----R-----H-----C----- 214  
Db 1526 QVYHPSRQPMAPNCEANGGQPCSHLCLINYNRTVSCACPHLMKHLKDNITTCVEPKFL 1585  
QY 215 -----I-----RE----- 217  
Db 1586 LYAROMBEIRGVDLAPYNYIISFTVPDINVTVDYDAREQRYWSDVTRTOAKRAFIN 1645

QY 218 G-G-----ODVPSNK--D-V-----TS-----LD----- 232  
Db 1646 GTGVETVVSADLP--NAHGLAVDWVSNRLFWTNTKQINVARLDGSKFNNAVQGLEQP 1704  
QY 233 -----W-----N--S-----EG--TLL--A--TG-----S--Y-- 245  
Db 1705 HGLVHVPLRGLKLYWTDGDNISMANMGDSNRTLFSQOKGPVGLAIDFPESKLYWISSGNH 1764  
QY 246 -----D-GF-----A-R-----I-----W-----T--K-DG----- 256  
Db 1765 TINRCLDGSGLVEIDAMRSQKATALAIMGDKLWADQVSEKMGTCRADGSGSVLR 1824  
QY 257 N--LA-----S-----T-----L--GO- 264  
Db 1825 NSTLWVHMKYVDESITQDHLKGTNPCSVNNGDCSQLCLPSETTRSCMCTAGISLRSQQ 1884  
QY 265 -----HKG-----P-----I--P-A-----L- 272  
Db 1885 ACEGVGSFLLYSVHEGIRGIPDPNDKSDALVPVSGTSLAVGIDFHAENDTIYVMDMGLS 1944  
QY 273 -----K-----W-----N--K-----KGN-----P--I--L-----S 283  
Db 1945 TISRAKRDQTRVEDVVTNGIGRVEGIAVDWIAWNIYWTQGFVIEVARLNGSFRYVIS 2004  
QY 284 AGVDK-----T-----T-----I-----I-- 292  
Db 2005 QGLDKPRAITVHPEKGYLFWTEWQYPRIERSLDGTERVVLVNSISWNGISVDYQDG 2064  
QY 293 --W-DA-----H----- 296  
Db 2065 KLYWCARDTKIERIDLETGENREVLLSSNMDFSVSFEDFIYWSDRTHANGSIKRG 2124  
QY 297 -----TG-----E-----AK-----OO----- 303  
Db 2125 KDNATSDVPLRTGIGVOLKDIKVFNRDRQGTNVCVAVANGCQOLCLYRGRGORACAH 2184  
QY 304 -----P----- 304  
Db 2185 GMLAEDGASCREVAGVLLYSERTILKSIHLSDBERNLAPVQPFEDPEHMKVIALAFDYR 2244  
QY 305 -----P-----F--H-----S-----A----- 309  
Db 2245 AGTSPGTNPNRIFPSDIHFQNIQIINDGSRITIVENVSGVEGLAYHRGMDTLWTSYTT 2304  
QY 310 -----P-AL-----D-----V-D-----W--O-----S 318  
Db 2305 STITRHTVDQTRPGAPERETVITMSGDHPRAEVLDECONLMFWTNNEQHPSIMRAALS 2364  
QY 319 --N-----N-----T--F--AS--C-----S----- 326  
Db 2365 GANVLTLEKIDITPGLAIDHRAEKLYFSDATLDKIERCEYDGSYHYVILKSEPHVPG 2424  
QY 327 -----TD-----M-----C-I 331  
Db 2425 LAVYGEHFWTDVRRVAVQRANKHGVSNMKLRLVDIPQOPMGIIVANDTNSCELSPCI 2484  
QY 332 -----HV--CKLG--OD-----R----- 340  
Db 2485 NNGGQDCLCLLTHQGHVNCSCR--GGRILQDLDLTCRAVNSSCRAQDEFECAENGECINFSLT 2543  
QY 341 -----P-----I--KT 344  
Db 2544 CDGVPCHKDSDEKPSYCNRRCKTFRQCSNGRCVSNMLWNCNGADDCGDSDEI PCNKT 2603  
QY 345 -----FQ----- 346  
Db 2604 ACGVEFRCKRGTGTCIGNSRQCNQVDCEDASDEMNGSATDCSSYFRLGVKGLFQPCERT 2663  
QY 347 -----G-----H 348  
Db 2664 SLCYAPSWVDGANDCGYSDERDCVGVKRPCLNYPACPSGRCPIMSWTCKEDDCEH 2723  
QY 349 --T--N-----EV-----N-A--I-K-W-----D-----PT-----G 361



Db 2724 GEDETHCNKFCSEAQFEQCNHRCISQWLCGSDGDDGDEAAHCEGKTCGSPSCPG 2783  
Qy 362 N-----L-----LAS-C--S--DD--M-----371  
Db 2784 THVCVPERWLCDGDKCAGADESIAAGCLYNSTCDDRFMCONQCIPKHFVCDHRDC 2843  
Qy 372 -----T-----Lk-----I--W-----S--MK 379  
Db 2844 ADGSDSPCEYPTCGSPSEFCANGRLSSRQWECDECHDQSDAPKPNHCTSPK 2903  
Qy 380 -----QD-----381  
Db 2904 CNASSQFLCSSGRVBAELLNQDCGSDSSDERGCHINECLSRKLSGCSQDCEDLKIGF 2963  
Qy 382 -----N-----CV-----H-----385  
Db 2964 KCRPGFRLKDDGRTCADVDECSFTTFCPSQRCINTHGSYKCLCVBGPYAPRGDPSCKA 3023  
Qy 386 -----D-----L-Q-----Q-----389  
Db 3024 VTDEEPFLIFANRYLRLKLDGNSYLLKQGLNNAVALDFDYREQMIYWDVTTOGSMI 3083  
Qy 390 -----H-----N-----K-----EI-----Y-TI-----397  
Db 3084 RRMHLGNSVQVLRGLTGLSNPDGLAVDWGGLNYWCDKGRDTIEVSKLNGAYRTLVLSSG 3143  
Qy 398 --K-----W-----S-----P-----T-----402  
Db 3144 LREPRALVVDVQNGYLYWTDGHDHSLIGRIGMDSSRSVIVDTKITWPNGLTLDYVTERI 3203  
Qy 403 -----G-----P-----T-K--H-----Q-----442  
Db 3204 YWADAREDI EFASLDGSRHVVLSQDIPHI FALTIFEDYVYWDNETKSNINRAHKTGT 3263  
Qy 407 N-----N-P-----N-ANL-MLAS-----A--S-F- 420  
Db 3264 NKTLTLLTLRPMDLHVFHALRQDPVNPHEPKVNGGCSNLCIL-SPGGGHKACPTNFY 3322  
Qy 421 --D-----S--T--V--R--LW--DV-----D-----R-----431  
Db 3323 LGSGRGTCVSNCTASQFVCKNDKCIPIFWMKCTEDDCGSDHSDPDPCEPKRPGQFQCS 3382  
Qy 432 -GTC-----IHT-L-----T-K--H-----Q-----442  
Db 3383 TGICTNPAFICDGNDCQNSDEANDCIHVCLPSQFKCTNTRNCIPIFCNGQDNCGDG 3442  
Qy 443 -----E-----P-VY-----S-----V-AF-----450  
Db 3443 EDERDCPEVTCAPNQFQCSITKRCIPRVWVCDRNDNCVDGSDPANCTQTCGVDFRCK 3502  
Qy 451 -S-----P-----DG-----R--Y-----LAS--G 460  
Db 3503 DSGRCIPARWKCDGEDDCGSDGDEPKECECDERTCEPYQFRCKNNRVCPGRWQCDYDNDG 3562  
Qy 461 -----S-F-----D-K-CV-----466  
Db 3563 DNSDEBSCTPRPCESEFSCANGRCIAGRWKGDHDCADGSDDEKDCPTRCDMDQFQCKS 3622  
Qy 467 -H-I--W-----NT-----471  
Db 3623 GHCIPLRWCDADACMDGSDDEACGTGVRTCPIDBFQCNNTLCKPLAWKCDGEDDCGN 3682  
Qy 472 -----QVCL-----H--476  
Db 3683 SDENPECARFVCPNRPFRCKNDKRCIWTGRQCDGTDNCGDGTDEBCEPPTAHTHCK 3742  
Qy 477 --YL-----479  
Db 3743 DKKEFLCNRQRLSSSLRCNMFDGSDGDEEDCSIDPKLTSCATNASICGDBARCVTE 3802  
Qy 480 --GQ-----V--L-----L-NL--G-----R--S--IC--491

Db 3803 KAAYCACRSGFHTVPQPGCQDINECLRFTGTCQLCNNTKGHLCSARNFMKTHNTCKA 3862  
Qy 492 -----LV-----493  
Db 3863 EGSEYQVLIADNDNEIRSLFPQPHPSAYEQAFQGDSESVRIDAMDVHVHAKGRVYVWTHWHTG 3922  
Qy 494 T-----LP-----H-----HL-----V-----VI 502  
Db 3923 TISYRSLPAPPTTSNRHRRQIDRGVTHLNISGLKMPRGIAIDWVAGNVYMTDSGRDVI 3982  
Qy 503 -----P-LVA--L-----IE-----L--LV-- 512  
Db 3983 EVAQMKGENKRTLISGMIDEPHAIVVDPLRGTMYSWDGWNHPKXIETAAMDGTTLRETLVQD 4042  
Qy 513 -----L-----K 514  
Db 4043 NIQWPTGLAVDYHNERLYWADAK 4065

RESULT 3  
US-08-469-658-52  
; Sequence 52, Application US/08469658  
; Patent No. 5917018  
; GENERAL INFORMATION:  
; APPLICANT: Th egeresen, Hans Christian  
; APPLICANT: Holtet, Thor Las  
; APPLICANT: Ezerodt, Michael  
; TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOLDING OF  
; TITLE OF SEQUENCES: 58  
; NUMBER OF SEQUENCES: 58  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Fish & Richardson P.C.  
; STREET: 225 Franklin Street  
; CITY: Boston  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02110-2804  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version  
; SOFTWARE: #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/469,658  
; FILING DATE: June 5, 1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/192,060  
; FILING DATE: February 4, 1994  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Paul T. Clark  
; REGISTRATION NUMBER: 30,162  
; REFERENCE/DOCKET NUMBER: 06363/002002  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 617 542 5070  
; TELEFAX: 617 542 8906  
; TELEX: 200154  
; INFORMATION FOR SEQ ID NO: 52:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4544 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-469-658-52

Query Match 70.2%; Score 2562.6; DB 2; Length 4544;  
Best Local Similarity 10.4%; Pred. No. 3.9e-32;  
Matches 414; Conservative 79; Mismatches 16; Indels 3474; Gaps 366;  
Qy 1 MS-----I-----SSDE-----V-----N--F-L-----11

Db 88 MSRLCNVQDCMDSDEGPHCRELOQNCRLGCOHCHVPTLDGPTCYCNSSFQLOADGKT 147  
QY 12 -----VY-----R-----Y-LQ-----E----- 18  
Db 148 KDFDECSVYGTCSQLCTNTDGSFICGCVGYLLQPDNRSCAKNEPVPDPVLLIANSQ 207  
QY 19 -----SG-----FS-----H-SA-----F 26  
Db 208 NILATYLSGAQVSTIPTSTRTQTMDFSYANETVCWVHVGDSAAQTOLKCARMPGLKGF 267  
QY 27 -----TFGIK-S-HISQ-----S-N-----I-----NG----- 41  
Db 268 VDEHTINISLSLHVEQMAIDLTGNFYFVDDDDIRIFVCNRNGDTCVTLLDLELYNPKG 327  
QY 42 -AL-----VP-----P-----A-A-48  
Db 328 IALDPAMGVFTDYGOIPKVERCDMDGQNRKTLVDSKIVFPHGITLTLVSLRVYWADAY 387  
QY 49 --LI-----STIQK-L-Q-Y--V-E-----A-EVS-I-----N 67  
Db 388 LDYIEVDYEGKGRQTIQ-GILIEHLYGLTVFENYLYATNSDANAAQKTSIVRVNRFN 446  
QY 68 --E-----D--GTL----- 72  
Db 447 STEYQVTVRDKGALHIYHORQPRVRSACENDQYKPGGCSIDICLLANSHKARTCRC 506  
QY 73 --F-----DG-----R-----PIE-----SL--81  
Db 507 RSGFSLGSDGSKCKPEHEFLVYKGRGPIIRGMMDGAKVDPDEHMIPIENLMPALDF 566  
QY 82 -----S-LI-----D-----AV-M-----PD--90  
Db 567 HAETGYFPADTYSYLGROKIDGTERETILKGIHNVEGVAVDMGDNLNWTDDGPCKT 626  
QY 91 V-V-----QTR-----QO-----97  
Db 627 ISVARLEKAAQTRKTLLEGKMTHPRAIVDPPLNGMWTWDWEDPKDSRRGRLERAWMDG 686  
QY 98 -----A-Y-R-----D-K-----102  
Db 687 SHRDIFVTSKVLWPNGLSLDIPAGRLYWDADFRIETILNGTRDKIVYEGPELNHAF 746  
QY 103 -----L-----AQOQAA-----109  
Db 747 GLCHGNYLFWTEYRSVYRLBERGVGAPPTVTLRSEPPPIFEIRMYDAQOQOVGTNK 806  
QY 110 -----AA-----A-A-----A-RAAQ-120  
Db 807 CRYNNGCSSILCLATPGSRQCAEADQVLDGVTCLANPSYVPPQCPQGFACANSRC 866  
QY 121 Q-----G-----S-----K-----N-G-----E 128  
Db 867 IOERWKCDGNDCLDNDSEAPALCHOHTCPSDRFKCENNRCPNRLWLCGDNDGNDSEDE 926  
QY 129 -N-T-A-----N-----G-E-----E--N 136  
Db 927 SNATCSARTCPNPQFSCASGRCTPISWTCDLDDCGDRSDESASCAYPCTCPFTQFTCNN 986  
QY 137 G-----A-H-----T-----I-----141  
Db 987 GRCININWRCDNDGNDSEAGCSHSCSSTQPKCNSGRCPHEHWTCDGNDGNDGYSDE 1046  
QY 142 --AN-N-----HTD-----M-----M-E-----V-D-152  
Db 1047 THANCTNQATRPFGGCTHDFQCRLDGLCPLRWCRDGDTCMDSDSKCEGVTHVCDP 1106  
QY 153 -----G-D-----V-----E-I--PP-----NKAV-----V-164  
Db 1107 SVKFGKDSARCTSKAWVCGDNDCEDNSENCESLACPPHPCANNTSVCLPPDKLC 1166  
QY 165 -----L-R-----G-H-----E-S-----V---172

Db 1167 DGNDCGDSDEGELCQCQSLNNGGSHNCVAPGEGIVCSCPLGMELGPDNHTCQIQSY 1226  
QY 173 -----F-I-CA-----W-----N-----178  
Db 1227 CAKHLKCSQKQDNKFSVKSCSYEGWVLEPDGSGSCSLDPFKPPIIFSNRHEIRRIDLHK 1286  
QY 179 -----P-----V-----S-----D-----LL--A-S-----186  
Db 1287 GDYSVLVPGRLNTIALDFHLSQSALYWTVDVEDKIYRGKLLDNGALTSFEWVIOYGLATP 1346  
QY 187 -----G-----S-----GD-----S-----T--192  
Db 1347 EGLADWIAAGNIYWVESNLDQIEVAKLDGTLRTLLAGDIEHPRAIALDPRDGILFTWDW 1406  
QY 193 -A--RI-----W-N----- 197  
Db 1407 DASLPRIASMSGAGRRTVHRETGGWPNGUTVLYLEKRIILWIDARSDAIYSARYDGS 1466  
QY 198 -----LS-----E-----NST-----SG-----S-TQ--L 209  
Db 1467 GHMEVLRGHEFLSHPPFAVTLYGGEVYWTDRTN-TLAKANKWTGHNVVQVQRTNTPFDL 1525  
QY 210 -V-----L-----R-----H-----C-----214  
Db 1526 QVYHPSRQPMAPNCPCEANGGQGPCSHLCLINYNRTVSCACPMLMKLHKDNTTCYEFKFL 1585  
QY 215 -----I-----RE----- 217  
Db 1586 LYARQMEIRGVLDADPYNYVISTVDPDINDTVLDYDAREQRYVSDVDTQAIKRAFIN 1645  
QY 218 G-G-----QDVPSNK-----D-V-----TS-----LD-----232  
Db 1646 GTGVEITVSADLP-NAHGLAVDWVSRNLFWTSYDTNKKQINVARLDGSPKNAVVGLEQP 1704  
QY 233 -----W-N-S-----EG-----TLL--A-TG-----S-Y-----245  
Db 1705 HGLVHPLRGKLYTQDGNISMANMDGSRNLTLLFSQKGPVGLAIDFPPEKLYWISSGNH 1764  
QY 246 -----D-GF--A-R-----I-----W-----T-K-DG-----256  
Db 1765 TINRCNLDGSGLEVIDAMRSQKATALAIMGKLMWADOVSEKMGTCSDAGSGSVLR 1824  
QY 257 N--LA-----HKG-----P-----I-F-A-----L--264  
Db 1825 NSTLWHEMKVYDESIOLDHKGTNPGSVNNGDCSQLCLPTSETTRSCMCTAGYSLRSQQ 1884  
QY 265 -----K-----K-----KGN-----F-I--L-----S 283  
Db 1885 ACEGVGSFLLYSVHEGIRGIPLDPNKSDALVPVSGTSLAVGIDFHAENDTIYVDMGLS 1944  
QY 273 -----K-----W-----N-----K-----KGN-----F-I--L-----S 283  
Db 1945 TISRAKRDQTRVEDVNTNGIRVEGIAVDWIAAGNIYWTQDGFVIEVARLNGSFYVVIS 2004  
QY 284 AGYDK-----T-----T-----T-----I-----I-----292  
Db 2005 QGLDKPRAITVHPKEGVLFWTEWQYPIERSRLDGTERRVVLNVSVISWENGISVDYQDG 2064  
QY 293 --W-DA----- 296  
Db 2065 KLYWCDAITDKIERIDLETGENREVVLSSNNMDMFSVSVFEDFYWSDRTHANGSIKRG 2124  
QY 297 -----TG-----E-----AK-----OQ-----303  
Db 2125 KONATDSVPLRTIGVQLDKIFVNRDRQKGTNVCAVANGCQOLCLYRGRGORACAH 2184  
QY 304 -----F----- 304  
Db 2185 GMLAEDGASCREYAGLYLSEYRILKSIHLSDERNLNAPVQPFDPHEMKNVIALAFDYR 2244  
QY 305 -----P--F--H-----S-----A-----309  
Db 2245 AGTSPGTNRIFFSDIHFHFGNIQOINDGSRRTITVENVGSVEGLAVHGRWDTLTYWTSYTT 2304

310 QY ---P-AL---D---V-D---W-Q---S 318  
311 |||  
2305 STITRHTVQTRGAFERETVITMSGDHPRAFLVDECNLMFTWNNQHPHSIMRAALS 2364  
312 |||  
319 -N---N---T---P---AS---C---S 326  
320 |||  
2365 GANVLTLIEKDINTPGLAIDHRAELKLYFSATLDKIERCEYDGSRYVILKSEPVHPFG 2424  
321 |||  
327 -TD---M---C-I 331  
328 |||  
2425 LAVYGEHIFWTDWVRRAVORANKHVSNMKLLRVDIPQPFMGIIAVANDTNSCELSFCRI 2484  
329 |||  
332 -HV---CKLG---QD---R--- 340  
333 |||  
2485 NNGGCQDLCLLTHQGHVNCSC-RGRILQDLTCAVNSCRAQDEFECANGBCINFSLT 2543  
334 |||  
341 -P---I---KT 344  
342 |||  
2544 CDGVPCHKDKSEKPSYCNRRCKTFRQCSNGRCVSNMLWCNAGDCCGDSDEIPCNKT 2603  
343 |||  
345 -FO--- 346  
346 |||  
2604 ACVGFEFRCDGTICGNSRCNQFVDCBDADEMNCSATDCSYFRLGVKVLFOPCERT 2663  
347 |||  
348 -G---H 348  
349 |||  
2664 SLCYAPSWCDGANDCGDYSDERDCPGVKRPRCLNYFACPSGRCLPMSWTCKEDDCEH 2723  
350 |||  
349 -T-N---EV---N-A-I-K-W---D---PT---G 361  
351 |||  
2724 GEDETHCNKFCSEAQCQNHRCISQWLCDGSDGDSDEAAHCEGKTCGSPSPFCP 2783  
352 |||  
362 N---L---LAS-C---S-DD---M--- 371  
353 |||  
2784 THVCVPERLWCDGDKDCADESIAGCLYNSTCDDREBFMCNQRCIPKHFVCDHRRDC 2843  
354 |||  
372 -T---LK---I---W---S-MK 379  
355 |||  
2844 ADGSDSEPEYTCGPSEFCANGCLSSRQWECGENDCHDSDEAPKNPHCTSPCHK 2903  
356 |||  
380 -QD--- 381  
357 |||  
2904 CNASSQFLCSGRVAEALLCNGQDDCGSDGSDRGCHINECLSRKLSGCSQDCDLKIGF 2963  
358 |||  
382 -CV---H--- 385  
359 |||  
2964 KCRCPGFLKDDGRTCADVDECSFTTFCPSQRCINTHGSYKCLCVBEGYAPRGDPSCKA 3023  
360 |||  
386 -D---L-Q---O--- 389  
361 |||  
3024 VTDEEPFLIFANRYLRKLNLDGNTLLKQGLNNAVALDPDYREQMIYWDVTTOGSMI 3083  
362 |||  
390 -H---N---K---EI---Y-TI--- 397  
363 |||  
3084 RRMHLNGSNVQVLRHTGLSNPDGLAVDWGNGLYWCDKGRDITIEVSKLNGAYRTLVSSG 3143  
364 |||  
398 -K---W---S---P---T--- 402  
365 |||  
3144 LREPRALVVDVQNGYLYWTDWGDHSLIGRIGMDGSSRSVIVDTKITWPNGLTLDYVTERI 3203  
366 |||  
403 -G---P--- 406  
367 |||  
3204 YWADAREYIEFASLDGSRHVVLSQDIPHI FALTLPEDYVYVWTDWETKSNRAHKTGT 3263  
368 |||  
407 N---N-P---N---ANL-MLAS---A-S-F 420  
369 |||  
3264 NKTLLISTLHRPMDLHFHALRPDPVNPHPCKVNGGCSNLCIL-SPGGGHKACAPTNY 3322  
370 |||  
421 -D---S-T---V---R---LW---DV---D---R--- 431  
371 |||  
3323 LGSDBGRTCVSNCTASQVCKNDKCIPIFWKNCDETDGDSDEPPDCBPKCRPGQFQCS 3382

432 -GIC---IHT-L---T-K-H---Q--- 442  
3383 TGICTNPAFICDGDNDQDSDEANCDIHVCLPSQFKCTNTRNCIPGFRNCQDNCGDG 3442  
443 -E---P-VY---S---V-AP--- 450  
3443 EDERDCPEVTCAPNQFQCSITKRCIPRVVWDRDNDVDSDEPANCTQMTQGVDFRCK 3502  
451 -S---P---DG---R---Y---LAS-G 460  
3503 DSGRCIPARWCKDGEDDCGSDGDEPKKECDERTCEPYQFRCKNNRCVPGRWQCDYNDG 3562  
461 -S-F---D-K-CV--- 466  
3563 DNSDEESCTPRPCSESEFSCANGRCIAGRWKCDGHDHCDGSDKDCOTPCRCMDQFOCKS 3622  
467 -H-I---W---NT--- 471  
3623 GHCIPLRWRCADACMOGSDDEACGTGVRTCPLDFEQCNNTLCKPLAWKCDGEDDCGN 3682  
472 -OVCL---H--- 476  
3683 SDENPEECARFVCPNRPFRCKNDRVCLWIGRCQDGTNCGDGTDEEDCEPPTAHTHCK 3742  
477 -YL---N--- 479  
3743 DKKEFLCRNORCLSSSLRCNMFDDCGSDGDEEDCSIDPKLTSCATNASICGDEARCVTE 3802  
480 -GO---V---L---L-NL-G---R---S-IC--- 491  
3803 KAAYCACRSQFHTVPGQCODINECLRFGTCSQLCNNTKGHLCSARNFMKTHNTCKA 3862  
492 -LY--- 493  
3863 EGSEYQVLIADNNEIRSLFPGPHSAYEQAFQGDSESVRIDAMDVHVKAGRVYTWHTG 3922  
494 T---LP---H---HL---V---VI 502  
3923 TISYRSLPPAAPPTTSNRHRRQIDRGVTHLNTISGLKMPRGIAIDWAGNVYWTDSGRDVI 3982  
503 -P-LVA-L---IE---L---LV--- 512  
3983 EVAQMKGENRKLISGMIDEPHAIIVDPLRGTYWSDWGNHPKPIETAAMDGLRLTLVQD 4042  
513 -L---K 514  
4043 NIQWPTGLAVDYHNERLYWADAK 4065

## RESULT 4

US-08-476-515A-84  
; Sequence 84, Application US/08476515A  
; Patent No. 6239270  
; GENERAL INFORMATION:  
; APPLICANT: Akerstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Raak, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjalms, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
; NUMBER OF SEQUENCES: 84  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Martin Savitzky  
; STREET: Rhone-Poulenc Rorer Inc., 500 Arcola Rd.;  
; STREET: 3C43,  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: Compaq PC  
 ; OPERATING SYSTEM: Windows 95  
 ; SOFTWARE: Word 7.0 (Patentin)  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/476,515A  
 ; FILING DATE: 07-JUN-1995  
 ; CLASSIFICATION: 435  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: US 08/344,836  
 ; FILING DATE: 23-NOV-1994  
 ; PRIOR APPLICATION DATA:  
 ; APPLICATION NUMBER: WO PCT/SE94/00483  
 ; FILING DATE: 24-MAY-1994  
 ; APPLICATION DATA:  
 ; APPLICATION NUMBER: SE 9301764-8  
 ; FILING DATE: 24-MAY-1993  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Savitzky, Martin  
 ; REGISTRATION NUMBER: 29,699  
 ; REFERENCE/DOCKET NUMBER: A1355D  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 610-454-3816  
 ; TELEFAX: 610-454-3808  
 ; INFORMATION FOR SEQ ID NO: 84:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 4654 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 ; US-08-476-515A-84

Query Match 70.1%; Score 2558.9; DB 3; Length 4654;

Best Local Similarity 9.5%; Pred. No. 5.2e-32;  
Matches 430; Conservative 60; Mismatches 18; Indels 4031; Gaps 371;

QY 1 M-----S-----I-----S-----S-----SDE- 7  
 DB 1 MDRGPAAVACTLLALVACLPASGQECDSAHFRCGSGHCIPADWRCDGTGKCSDDADEI 60  
 QY 8 ---VN---F---L---V-----12  
 DB 61 GCAVVTCCQGYFKQSEBGQCIIPSSVWCDQDQDGDGSDRQDCSQSTCSSHQITCSNGQC 120  
 QY 13 ---YR---Y---L---Q-----E---S 19  
 DB 121 IPSEYRCDHVRDCPDGADENDCOYPTCEQLTCDNGACYNYSQRCDWKVDCRDSSEINCT 180  
 QY 20 ---G-FS-----H-----S-A-F-T---F---G-I---30  
 DB 181 EICLHNEFCGNECIPRAVYCDHNDCCQDGSDEHACNYPTCGYQPTCPSGRCIYQNVV 240  
 QY 31 ---K-----S-----H-----IS-----Q-SNI 39  
 DB 241 CDGEDCKDNGDEDCGSGPHVHKCSPREWSCPSGRCISIVKVCDDGILDGCPGREDEN- 299  
 QY 40 N---G-----AL---VP-----PAALI---S-----I-I---53  
 DB 300 NTSTGKYCSMTLCSALNCQYCHETPYGGACFCPPGGYIINHNSRTCEVFDDCQIWGICD 359  
 QY 54 QK---G-L-L-QY-----VE 61  
 DB 360 QKCESRPRHLCHCEGYILLERQYCKANDSPGEASIFNSGRDLLIGDIGHRSFRLIVE 419  
 QY 62 ---A-----E-----VSI-----N 67  
 DB 420 SQNRGAVGVAFVHVLQRVFTDTQNVKFSVDINGLNIQIEVLNVSVETPENLAVDWVN 479  
 QY 68 ---E-----DG---TL-----72  
 DB 480 KIYLVETKVRNIDWNLWDGSRVTLITENLHPRGIAVDPTGYLPFSDWESLSGPKLE 539  
 QY 73 --F-DG-R-----P-----IE-----S 80

DB 540 RAFMDGSRKDLVKTKLGPAGVTLDMSKRVVWDSRFDYIETVTYDGIQRTKVHGG 599  
 QY 81 L-----SL-----I---D---AVM-----88  
 DB 600 LIPHPFGVSLFEGQVFTDWTQKAVLKANKFTETNPQVYQASLRPQVGVTVVHSLRQPYA 659  
 QY 89 --P-D-----V-V-----Q-T-R-----Q-QAY 99  
 DB 660 TNPCKDNNGGCEQVCLSHRTDNDGLGFRCKTFGFGOLDTDERHCAVQNFILFSSQVAI 719  
 QY 100 R-----D-K-----L-AQ-----Q-Q---A 108  
 DB 720 RGIPFTLSTQEDVMVPVSGNPFVFCIDFDAQDSTIFFSDMSKHMIFKQKIDGTGREILA 779  
 QY 109 A-----A-----A-----A-----A---112  
 DB 780 ANRVENVESLAFDWISKNLYWTDSDHYKISVMRLADKTRRTVVQYLNNPRSVVHPFAGY 839  
 QY 113 -----A-----A-----A-AA-----S---119  
 DB 840 LFTDWRFRPAKIMRAWSDGSHLLPVINTTLGWPNGLAIDWAASRLYVWDAYFDKIEHSTF 899  
 QY 120 -----Q-----120  
 DB 900 DGLDRRLGHIEQMTHPFGLAIFGEHLFFTDWRLGAIIRVRKADGGEMTVIRSGIAVILH 959  
 QY 121 -----Q-GS-A-----KNG-----EN-----TA-N---G---133  
 DB 960 LKSYDNIQTGSMACNQTHPNQDCHSHFCFPVFNFORVCGCPYGMRLASNHLTCEGDPTN 1019  
 QY 134 E-E-E-----N-----GA-H-----T-----140  
 DB 1020 EPTEQGLFSPCKNGRCVPNYYLDCGVDDCHDNDSDQLCGTLNNTCSSAFTCHGEC 1079  
 QY 141 I-A-----N-----N-H-----T-DMME-----V-----DG-D 154  
 DB 1080 IFAHWCDCXKNDVCGSDSHNCPTTHAPASCLDTQYTCNHCQICSKWVCDDTDNDGCGSD 1139  
 QY 155 ---VEI---P-----PN-----K-A---162  
 DB 1140 EKNCNSTETCQPSQFNCNHRCIDLSFVCDGDKDCVDSDEVCVLNCTASQPKCASGDK 1199  
 QY 163 ---V-----VL-----R-G-H-----E-----169  
 DB 1200 CIGVTNRCDGVDCSDNSDEAGCPTRPGWCHSDEFQCEQEDGICIFNFWECGHDPCLYG 1259  
 QY 170 S-E---V-----F---IC---AW---N-----P-----179  
 DB 1260 SDEHNACVPTCPSSYFHCNGNCIHRAMLCDRDNDGDMSDKDCPTQFRCPSMQWQC 1319  
 QY 180 -----V-----SD-----182  
 DB 1320 LGHNICVNLVWCDGIFDCPNGTDESPLCNGNSCSDFNCGCTHECQVEPFGAKCLCPG 1379  
 QY 183 LLA-----S-----GS-----G-----189  
 DB 1380 LLANDSKTEDIDECIDILGSCSOHCYNMRGSPFCSDTGYMLESDBRTCKVTASESLLL 1439  
 QY 190 -----DS-T-----A-----RI-----W---N---197  
 DB 1440 VASQNKIADSVTSQVHNISLVENGSYIVAVDFDSISGRIFWSDATQGTWAFQNGTD 1499  
 QY 198 -----LSE-----N-S---T---S---GS-----T-Q-LVL 211  
 DB 1500 RRWVFSSIIETIAIDWVGRNLYWTYALETIEVSKIDGSHRTVLISKNLTPRGLAL 1559  
 QY 212 --R-----H-----213  
 DB 1560 DPRMNEHLLFWSDGWHHPRIERASMDGSMRTVIVQDKIFWPCGLTIDYPNRLLYFMDSYL 1619  
 QY 214 ---C-----IR-----E-----GG-----219

Db 1620 DYMFCDYNGHRRQVIASDLIRHPYALTFLFEDSVYVWTDTRATRRVMRANKWHGNGSVV 1679  
Qy 220 -----Q----- 220  
Db 1680 MYNIQWPLGIVAVHPKQNSVNPFCASRCSHLLSSQGFPHFYSCVCPGWSLSPLLN 1739  
Qy 221 -----V-----P-----SN-----K-----DV----- 228  
Db 1740 CLRDDQPLITVRQHIIIFIGISLNEPVKSDAMVPFIAGIQLDVEFDDAEQIYVWENPG 1799  
Qy 229 -----T-----S-----LDW-----N-----S-----E-----G-----T 238  
Db 1800 EHRVKTGNTVRVFAISVMWGPMSNLALDWISRNLYSTNPTQTSIEVLTLLHGDIRYRKT 1859  
Qy 239 LLA-----T-----G-----S-----Y-----DG-----F----- 248  
Db 1860 LIANDGTALGVGPIGITVDPARGKLYWSDQGTDSGVPAKIASANMDGTSVKTLFTGNLE 1919  
Qy 249 -----A-----R-----I-----W-----T----- 253  
Db 1920 HLECVTLIDIEQKLYAVTGRGVIERNVGDTRMILVHQLSHPWGIAVHDSFLYTTDEQ 1979  
Qy 254 -----K-----D-----G-----N-----J-----ASTL-----G-----Q-----H-----K----- 266  
Db 1980 YEVIERYDKATGANKIIVLRDNVPLNRLGLQVYHRRNAEASSNGSCNNMNACQOICLPVPGG 2039  
Qy 267 -----G-----P-----P-----I-----PALK----- 273  
Db 2040 LFSCACATGFLPNDRSCFPYNSFIVVMSLSAIRGFSLSLSDHSETMVPVAGQGRNALH 2099  
Qy 274 -----W-----N-----K-----K-----GN----- 279  
Db 2100 VDQVSSGFYWCDFSSVASDANAIIRIKPDGSSLMNIVTHGICENGVRGIAVDWVAGNL 2159  
Qy 280 -----F----- 280  
Db 2160 YFTNAFVSETLIEVLINTTYRRVLLKVTVDMPRHVVDPKRYLEFWADYQRPKIERSF 2219  
Qy 281 -----I-----L-----S-----A-----G----- 285  
Db 2220 LDCNRTVLVSEGI VTPRGLAVDRSDGYVTVWDDSLDIIARINGENSEVIRYGSRYPT 2279  
Qy 286 -----VD-----K-----T-----TII-----W-----D----- 294  
Db 2280 PYGITVPFENSIIWDRNLKXIFQASKEPENTEPPTVIRDNIINLRDVTIFDKQVQPRSPA 2339  
Qy 295 -----A-----HT-----G-----E-----A----- 300  
Db 2340 EVNNPCLENNCGCSHLCFALPGLHTPKCDCAFGLQSDGKNCAISTENFLIFALSNSLR 2399  
Qy 301 -----K-----Q-----O-----F-----P----- 305  
Db 2400 SLHLDPENHSPPQTINVERTVMSLDYSDVSDRIYFTQNLASGVQISYATLSSGHTPT 2459  
Qy 306 -----F----- 306  
Db 2460 VIASGIGTAGIAFDWITRIIYSDYLNQMINSMAGDSNRTVIARVPKRAIVLDPCCQ 2519  
Qy 307 -----H-----S-----AP-----ALD----- 313  
Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSSLVMPSGLTLDYEEDLLYVWDASLQRIERST 2579  
Qy 314 -----VD-----W-----Q-----S-----N----- 319  
Db 2580 LTGVDREVINAAVHAFGLTLYGQIYWTDLTYQRIYRANKYDGSQGIAMTTLNLSQPRG 2639  
Qy 320 -----NT-----F----- 322  
Db 2640 INTVVRNQKQCNPCPEQFNGGCSHICAPGNGAEQCPHEGWNWYLANNRKHCIVDNGER 2699  
Qy 323 -----AS-----CS-----TD-----M-----C-----TH-----V----- 333  
Db 2700 CGASSFTCSNGRCISBEWKCDNDNDGSDGSESVICALHTCSPTAFTCANGRCVQYSYR 2759

Qy 334 CKL-----G-----QD-----R-----P-----I----- 342  
Db 2760 CDYNDGCGSDEAGCLFRDCNATTFMCMNRRCIPREFICNGVDNCHDNTSDKNCPCD 2819  
Qy 343 -----K-----T----- 346  
Db 2820 RTCQSGYTKCHNSNICIPRVYLCGDGNDGSDNENPTYCTTHTCSSSEFQCASGRICIPQ 2879  
Qy 347 -----GH-----T----- 349  
Db 2880 HWYCDQETCFDASDPASCGHSERTCLADEFKDGRGRCIPSEWICDGDNDGSDMSDEDK 2939  
Qy 350 -----N-----E-----VNA-----IK-----W----- 357  
Db 2940 RHOCQONCSDSFELCVNDRPPDRRCIPQSWVCDGVDCTDGYDENQNCRTTCTCSENEPT 2999  
Qy 358 ----- 357  
Db 3000 CGYGLCIPKIFRCDRHNDGSDYSDERGCLYQTCQNOFTCQNGRCISKTFVCDENDDCGD 3059  
Qy 358 -----D-----PT-----G-----N-----L----- 363  
Db 3060 GSDELMHLCHTPEPTCPHFHKCDNGRCITEMMKLCNHLDDCLDSDKGCINECHDPSI 3119  
Qy 364 -----LAS-----CS-----D-----D-----M----- 371  
Db 3120 SGCDHNCTDLTISFYCSRGYKLMDSKRTCDVIDECTEMPFVCSQKCNVIGSYICKCA 3179  
Qy 372 ----- 371  
Db 3180 PGLYREPDGKTQRONSIBPYLIFSNRYLRNLTDIGYFYSLILEGLDNVVALDFDRVEK 3239  
Qy 372 -----T-----L-----K-----I-----W-----S----- 377  
Db 3240 RLYWIDTQRQVIERMFLNKTNETIINHLPAEASLAVDMVSRKLYWLARDLGLFVSXL 3299  
Qy 378 -----M-----KQ-----D-----N-----C-----VHDLQ-----H-----NK----- 392  
Db 3300 NGHRRMLAQHCVDANNTFCFDPNPRGLAH--POGYLYWADMGHRAIYGRVGMGTNKS 3357  
Qy 393 -----E-----I-----YT-----IK----- 398  
Db 3358 VIISTKLEWPNGITIDYTNDLLYWAADAHLYIEYSDLEGHHRHTVYDGLPHFPFAITFE 3417  
Qy 399 -----W-----S-----P-----T-----G----- 403  
Db 3418 DTIYWTDMNTRTVEKNGKYDGSNRQTLVNTTHRPFDIHVYHPYRQPIVSNPCGTNNGCS 3477  
Qy 404 -----PG-----T-----NN----- 408  
Db 3478 HLCLIKPGKGFTCECPDDFRTLQLSGTYCMPMCSSTOFLCANNKRCIPHWKCDGQKD 3537  
Qy 409 -----P-----NA-----N-----LML-----AS 417  
Db 3538 CSDGSDALALCPORFCLQFQCGDCNCTSPOTLCNAHQNCPCDGSDEDL--LCENHHCD 3596  
Qy 418 -----A-----S-----F-----DS-----T-----VR-----L-----W-----DV 429  
Db 3597 NEWQCANKRCIPESWQCDTFNDCEDNDSBDSHSCASRTCRPGQFRANGRCIPQAWKCDV 3656  
Qy 430 -----DRG-----I-----CI-----H-----T-----L----- 438  
Db 3657 DNDGSDHSEPIEEECSSAHLCDNFTEFSCKTNYRCIPKAVCNVGVDDCDRDNDEQCEE 3716  
Qy 439 -----T-----K-----H-----Q-----E-----P-----V----- 445  
Db 3717 RTCHPVGDFRCNMHHCIPLRWQCDQNDGNDSDENECAPRECTESEFRVCNQCIPSRW 3776  
Qy 446 -----Y-----S-----V-----A-----F 450  
Db 3777 ICDHYNDGSDNSDERDCMRTCHPEYFOCTSGHCVHSELKCDGSADCLDASDEADCPTRF 3836

```
QY 451 SPGRYL-AS-----GS-----F----- 462
Db 3837 -PDGAYCATMFECKNHCVCIPYWKCDGDDCGDSEHLCLDVPKSPNFRCDNR 3895
QY 463 -----D-----K-----CV-H-I----- 468
Db 3896 CIYSHEVCNGVDCGDTDETEHCKRPTKPKTEYKCGNGHCIPHNDVNCDDADDG 3955
QY 469 W-----N-TQ-----V-----CL----- 475
Db 3956 WSDLGCKGKERTCAENICEONCTQNEGFCCTAGFETNVFRTSCLDINECEQFG 4015
QY 476 -----H----- 476
Db 4016 TPOHCRNTKGSVEVCADGFTMSDRPKRCAEGSSPLLLPDNVRIRKYNLSRFS 4075
QY 477 -YL-----N-----G----- 480
Db 4076 EYLQDEEYIQANDYDWDPKDGLSVVYTYVRGSRFGAKRAVIPNFESGRNVLQEV 4135
QY 481 -----Q-----V-----L-----N----- 485
Db 4136 LKLKYNMOPDGIADVWGRHYWSDVKNKRIEVAKLDGRYKWLITDLDQPAIAVNP 4195
QY 486 LG-----R-----SI----- 490
Db 4196 LGLMFTWKGKPKIESAWNGEDRNILVFDLGNWPTGLSIDYNDRIYWSDFKEDI 4255
QY 491 ----- 490
Db 4256 IKYDGTDRRTAKEMNPYSIDIPEDQLYWISKEGBWKQKFGQKKEKTLVNPWL 4315
QY 491 -----CL-----Y-----T----- 494
Db 4316 QVRIFHOLRYNKSVPNLCKQICSHLCLLRPGYSACPOGSSFFIEGSTTECDAA 4375
QY 495 LP-----H----- 497
Db 4376 LPPRCRCHGNCYFDETDLPKCKPSGYTKYCEMAFSGISPGTTAVAVALLTILL 4435
QY 498 -----H-----LV-----V-----I-----P 503
Db 4436 IGALAIAGFYHRRGTGSLPALPKPLSLSLVKPSENGVTFRSGADLNMDIGVSG 4495
QY 504 -----LVA-----LI-E-LL-----VLK 514
Db 4496 ETAIDRSMASBEDFVMEKGQPIIFENPMYSARDSAV-K 4533
```

## RESULT 5

```
US-08-652-877-84
; Sequence 84, Application US/08652877
; Patent No. 6187548
; GENERAL INFORMATION:
; APPLICANT: Akerstrom, Goran
; APPLICANT: Juhlin, Claes
; APPLICANT: Rask, Lars
; APPLICANT: Crumley, Gregg R.
; APPLICANT: Morse, Clarence C.
; APPLICANT: Murray, Edward M.
; APPLICANT: Hjaln, Goran
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Rhone-Poulenc Rorer Inc.
; STREET: 500 Arcola Rd., 3C43
; CITY: Collegeville
; STATE: PA USA
; COUNTRY: USA
; ZIP: 19426-0107
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
;
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.5.1
; SOFTWARE: Word 6.0 (Patentin)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/652,877
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION NUMBER: PCT/US95/15203
; FILING DATE: 22-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/344,836
; FILING DATE: 23-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/487,314
; FILING DATE: 07-JUNE-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Savitzky, Martin
; REGISTRATION NUMBER: 29,699
; REFERENCE/DOCKET NUMBER: A1355E-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-454-3816
; TELEFAX: 610-454-3808
; INFORMATION FOR SEQ ID NO: 84:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4655 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-652-877-84
```

```
Query Match 70.1%; Score 2558.8; DB 3; Length 4655;
Best Local Similarity 9.5%; Pred. No. 5.3e-32;
Matches 430; Conservative 60; Mismatches 18; Indels 4032; Gaps 371;

QY 1 M-----S-----I-----S--SDE- 7
Db 1 MDKGPAAVACTLLALVACLAPASGQECDSAHFRGSGHCHIPADWRCDTGKDCSDDAEI 60
QY 8 ---VN---F---L---V----- 12
Db 61 GCNAVTCQOQYFKQSEGQCIPSSWVCDDQDQDCDDSDERQDCSQSCSSHQITCSNGQC 120
QY 13 ---YR-----Y---L---Q-----E--S 19
Db 121 IPSEYRCDHVRDCPDGADENDCOYPTCEQLTCDNGACYNTSQKCDWKVDCRSDSDEIN 180
QY 20 ---G-FS-----H---S--A--F-T---F--G-I----- 30
Db 181 EICLHNEFCGNGECIPRAYVVDHNDQCGSDGDEHACNYPTCGGYQFTCPSGRCIQNVV 240
QY 31 ---K-----S---H-----IS-----Q--SNI 39
Db 241 CDGEDCKNGDEDCGSGPHDVHKCSPREMSCPSGRCISIVKVCDDGILDCPREDEN- 299
QY 40 N---G-----AL-----VP-----PAALI-----S-----I-I-- 53
Db 300 NTSTGKYCSMTLCSALNCQYQCHETPYGGACFCPPGGYIINHNDSTRCTVEFDQCIW 359
QY 54 QK-----G-L--QY-----VE 61
Db 360 QKCESRPGRHLCHEEGYILRGQYCKANDSFGESAIIFSNGRDLLIGDIHGRSFRILVE 419
QY 62 ---A-----E---VSI-----N 67
Db 420 SQNRGAVGVAFHYHLQRFVFTVQNKVFSVDINGLNIQVNLVSVETENLAVDWNN 479
QY 68 ---E-----DG--TL----- 72
Db 480 KIYLVETKVNRIKDWNLDSYRVTLITENLHGRGIAVDPTVGYLFFSDWESLSGEPKLE 539
QY 73 --F-DG--R-----P-----IE-----S 80
```

Db 540 RAFMGDSNRKDLVKTGLGNPAGVTLDMISKRVYVWDSRFDYIETVTYDGIQRTVVHGG 599  
Qy 81 L-----SL-----I-----D-----AVM----- 88  
Db 600 LIHPFGVSLPEGQVFTDWTKAVLKANKFETNPQVYQASLRPGVTVYHSLRQPYA 659  
Qy 89 -P--D-----V--V-----Q--T--R-----Q-----Q--AY 99  
Db 660 TNPCKONGGCEQVCLSHRDTNDGLGRCKCTFGFOLDTDERHCIAVQNFIFSSQVAI 719  
Qy 100 R-----D--K-----L--AQ-----Q-----Q--A 108  
Db 720 RGIPFTLSTQEDVMVPVSGNPSFFVGIDFDAQDSTIFFSDMSKMFQKIDGTGREILA 779  
Qy 109 A-----A-----A-----A-----A-----A-- 112  
Db 780 ANRVENVESLAFDWISKNLYWTDSHYKSISVMRLADKTRTVVQYLLNPRSVVHPFAGY 839  
Qy 113 -----A-----A-----A-----A-----A-----S-- 119  
Db 840 LFPTDWRPAKIMRAWSGDGSHLLPVINTILGWPNGLAIDWAASRLYVWDAYFDKIEHSTF 899  
Qy 120 -----Q-----Q-----Q-----Q-----Q-----T-- 120  
Db 900 DGLDRRLGHIEQWTHPFGLAIFGEHLFFTDWRLGAIIRVKADGEMTVIRSGIAYILH 959  
Qy 121 -----Q--GS--A-----KNG-----EN-----TA--N-----G-- 133  
Db 960 LKSYDVNIQTGSNACNPTHPNGDCSHFCPPVFNQFVQVCGPYGMRLASHNLTCEGDPTN 1019  
Qy 134 E-----E-----N-----GA--H-----T-----T----- 140  
Db 1020 EPTEQCLGFFPKNGRCVPNYVLCGVDDCHNDSDEQLCGTLNNTCSSAFTCGHGE 1079  
Qy 141 I--A-----N-----H-----T--DME-----V-----DG--D 154  
Db 1080 IPAHRCDKRENDVCGSDENHCPHAPASCLDTQYTCNHNQICSKNWCVDTDNDGCGSD 1139  
Qy 155 -----VEI--P-----PN-----R-----K--A----- 162  
Db 1140 EKNCTETCOPQFNCNHRICDLSFVCGDKDCVDGDEVCVLCNCTASQFKCASGDK 1199  
Qy 163 -----V-----VL-----R--G--H-----E----- 169  
Db 1200 CIGVTRKCGVPCSDNSDEAGCPTRPFGMCHSDEFQCSGDGICIPNFEWCDGHPDCLYG 1259  
Qy 170 S--E-----V-----F-----IC-----AW-----N-----P----- 179  
Db 1260 SDEHNACVPKTCPPSYFHCNNGNCIHRAWLCDRDNDGDMSDKDCPTOPFRCPNQWQC 1319  
Qy 180 -----V-----V-----SD----- 182  
Db 1320 LGHNICVNLVCGDGFDCPNGTDESPLNGNSCSDFNGGCTHECVQEPFGAKCLPLGF 1379  
Qy 183 LLA-----S-----S-----G----- 189  
Db 1380 LLANDSKTCEDIDECILGSCSHQCYNMRGSPRSCDGTGYMLESDRGTCKVTASESILL 1439  
Qy 190 -----DS--T-----A-----RI-----W-----N----- 197  
Db 1440 VASQNKIIADSVTSQVHNISYLVENGSIYVAVDFDSISGRFWSDATQGTWSAFQNGTD 1499  
Qy 198 -----LSB-----N-----S-----T-----S-----GS-----T--Q--LVL 211  
Db 1500 RRVFDSIIETIATIDAVGRNLYWTDYALETIEVSKIDGSHRTVLIISKNLTNPRGLAL 1559  
Qy 212 --R-----H----- 213  
Db 1560 DPRMNEHLLFWSDGHHPRIERASMDGSMRTVIVQDKIFWPCGLTIDYPNRLLYFWD 1619  
Qy 214 -----C-----IR-----E-----GG----- 219  
Db 1620 DYMDFCDYNGHRRQVIASDLIIRHPYALTFLFEDSVYVTDTRATRRVRANKWHGQNSVV 1679

Qy 220 -----Q----- 220  
Db 1680 MYNIQWPLGIVAVHPKQPNVNPCAFSCRSCHLCLSSQGFPHFYSCVPCSGWSLSPDLLN 1739  
Qy 221 -----D-----V-----P-----SN--K-----DV----- 228  
Db 1740 CLRDDQFLITVROHIIIFGISLNPEVKSDAMVPIAGIQLNGLDVEFPDABEQYIYWVENPG 1799  
Qy 229 -----T-----T-----S-----LDM-----N-----S--E-----G-----T 238  
Db 1800 EHRVKTGDNRTVFASISMWGSPMMLALDWMISRNLYSTNPTQSIQEVLTGHDIRYRKT 1859  
Qy 239 LLA-----T-----G--S-----Y-----R--I-----W-----T----- 248  
Db 1860 LIANDGTALGVGPIGIVTDPARGKLYWSDQGTDSGPAPAKIASANMDGTSVKTLFTGNLE 1919  
Qy 249 -----A-----A-----R--I-----W-----T----- 253  
Db 1920 HLECVTLIDIEEQKLYWAVTGRGVIERNVDGTMILVHOLSHPWGIAVHDSFLYTTDEQ 1979  
Qy 254 -----K--D--G--N-----L-----ASTL--G--O--H-----K----- 266  
Db 1980 YEVIKATGANKIVLRDNVNLRLQVYHRRNAESNGCSNNMNACQICLPVPGG 2039  
Qy 267 -----G-----P-----I--FALK----- 273  
Db 2040 LFSCACATGKLPNDNRSCSPYNSPIVVSMLSAIRGFSLESDHSETWVPVAGQGRNALH 2099  
Qy 274 -----W-----N-----K-----K-----GN-- 279  
Db 2100 VDVSSGFIYWCDFSSVASDNRIRIKPDGSSLMNIVTHGIGENGVRGIAVDWVAGNL 2159  
Qy 280 ----- 280  
Db 2160 YFTNAFVSETLIEVLRIINTYRRVLLKVTVDMPRHIVVDPKRYLFWADYQORPKIERSF 2219  
Qy 281 -----I-----L-----S-----A-----G----- 285  
Db 2220 LDCNRTVLVSEGIPTPRGLAVDRSDGYVYVWDDSLDIATIRINGENSEVIRYGSRYPT 2279  
Qy 286 -----VD-----K-----T-----TII-----W-----D----- 294  
Db 2280 PYGITVFENSIIVWDRNLKKIFQASKEPENTEPPTVIRDININWLRDVTIFDKQVQPRSPA 2339  
Qy 295 -----A-----HT-----G-----E-----A----- 300  
Db 2340 EYNNPFCLENNGCCHLCPALPGLHTPKCDCAFGLTQSDGKNCAISTENFLIFALNSLR 2399  
Qy 301 -----K-----Q-----Q--F-----P----- 305  
Db 2400 SLHLDPENHSPPQTINVERTVMSLDYDSVSDRIYFTQNLASGVQISYATLSSGIHTPT 2459  
Qy 306 -----F----- 306  
Db 2460 VIASGIGTAGIAFDWITRRIYVDYLNQMINSMABDGSNRTVIARVPKPRAIVLDPQCG 2519  
Qy 307 -----H-----S--AP--ALD----- 313  
Db 2520 YLYWADWDTHAKIERATLGNFRPVIWSSLVMPSGLTLDYEBEDLLYWVDASLQRIERST 2579  
Qy 314 -----VD-----W-----Q-----S-----N----- 319  
Db 2580 LTGVDBREVINAHVAFGLTYQYIYWTDLTYQRIYRANKYDGSQIANTNLLSQPRG 2639  
Qy 320 -NT-----F----- 322  
Db 2640 INTVWKNQKQCNNPCEQFNGGCHICAPGPNAGECQCPHEGWNWYLANNRKHCIVDNGER 2699  
Qy 323 --AS--CS-----TD--M--C--IH-----V----- 333  
Db 2700 CGASSPTCSNGRCISBEWKCDNDGCGSDMESVICALHTCSPTAFTCANGRCVQYSYR 2759

334 CKL-----G--QD-----R-P--I-----342  
2760 CDYNDGDSGDEAGLFRDCNATTFMCMNRRCIPREFICNGVDNCHDNNTSDEKNCDD 2819  
343 -----K-----T-----FQ-----346  
2820 RTCQGYTKCHNSNICIPRYLVCDGNDGDSNENPTYCTTHTCSSSEFQACASGRICIPQ 2879  
347 -----GH--T-----349  
2880 HWYCDQBTDFCDASDEPASCCHERTCLADEFKDGGRCIPSEWICDGDNDGDSDEDK 2939  
350 -----N--E--VNA-----IK--W-----357  
2940 RHOCQNCQSDSEFLCVNDRPRRCIPQSVWCVGDVDCDGYDENQNCRTTRTCSENEFT 2999  
358 -----357  
3000 CGYGLCIPKIFCRDRHNDGDSYDERGCLYQTCQOQFTQNGRCISKTFVCDENDCGD 3059  
358 --D-----PT-----G-----N-----L-----363  
3060 GSDLMHLCHTPEPTCPPEBFKCDNGRCIEMMKLCNHLDDCLDNDSEKKGCGINECHDPSI 3119  
364 -----LAS--CS-----D-D--M-----371  
3120 SGCHNCTDTLTSPYCSRGYKLMKDKRTCDIDECTEMPFVCSQKCNVIGSYICKCA 3179  
372 -----371  
3180 PGYLRPDGKTRQNSNIEPVLIFSRYLNLNTIDGYFYSLILEGLDNLVALLDFRVEK 3239  
372 -----T-----L--K--I-----W-----S--377  
3240 RLYWIDTQROVIERMFLNKTNETIINHRLFAESLAVDWVRKLYLWDLARLDGLFVSD 3299  
378 -----M-KO--D--N--C-----VHDLOQ-----H-----NK--392  
3300 NGHRRMLAQHCVDANNTFCDNPRGLAH--POGYLYWADWGHRAIYGRVGMGTNKS 3357  
393 -----E-----I-YT-----IK-----398  
3358 VIISTKLEWNGITIDYNDLLYWADAHGYEYSDLEGHHRHTVYDGLPHPPAIFTE 3417  
399 -----W-----S-P--T--G--403  
3418 DTIYTDWNTREKNGKIDGNSNRQTLVNTTHRPFDIHVYHPYQPIVSNPCGNTNGGCS 3477  
404 -----PG-----T-----NN-----408  
3478 HLCLIKPGGKFTCECPDDFTLQISGTYCMPCSSSTQFLCANNEKCIPIWKKCDGQKD 3537  
409 -----P-----NA--N-----LML-----AS 417  
3538 CSDGDELALCPORFRLGQFCSDGNGTSQTLNHAQNCDDGSDERL--LCENHHCHDS 3596  
418 -----A-----S-----F-----DS-----T-----VR-----L-----W--DV 429  
3597 NEWQCANRKCIPESWQCDTFNDCEDNDESDSHCASRTCRPGQFRANGRCIPOAWKCDV 3656  
430 --DRG-----I--CI--H-----T-L-----438  
3657 DNDGDHSDPEIEBCMSAHLCDNFTEPSCNTNYRCIPKWAVCNGVDGCDNSDEQCEE 3716  
439 -T-----K--H-----Q-----E--P-----V-----445  
3717 RTCHPVGDFRCKNHHCIPLRWQCDQNDGDSNDEENCAPRECTESEFRVCNOQCIPSRW 3776  
446 -----Y-----S--V-----A-----F 450  
3777 ICDHYNDGDSNDESDERCMRTCHPEYFOCTSGHCVHSELKCDGADCLDASDEADCTRF 3836  
451 SPDGRYL--AS-----GS-----F-----462

3837 -PDGAYCATWFECKNHCVCIPYWKCDGDDDCGDSDEELHCLDVPNCSPNRFRCNNR 3895  
463 -----D-----K-----CV-H--I-----468  
3896 CIYSHEVCNGVDDCGDGTDETEHCRCRTPKCTPEYKCGNGHCIPHDNVCCDADDCGD 3955  
469 W-----N--TQ-----V-----CL-----475  
3956 WSDLGCKNGKERTCAENICEQCTQINEGGFICSTAGTAGFETNVFDRISCLDINECEQFG 4015  
476 -----H-----476  
4016 TCPQCRNTKSGYECVADGFTMSDRPGKRCABEGSSPLLLPDNVIRIRKYNLSRFS 4075  
477 -YL-----480  
4076 EYLODEEYIAVDYDWDPKDGLSVVYTVRGEGRFGAIKRAYIPNFGSRNLLVQEVVD 4135  
481 -----Q-----V-----L-----L-----N--485  
4136 LKLKYMQPDGIADVWGRHIYWSVKNKRIEVAKLGDGRYKWLSTDLDPAAIAVNEK 4195  
486 LG-----R-----SI-----490  
4196 LGLMFWTDWKEPKIESAMWNGEDRNILVFDLWPTGLSIDYLNNDRIYWSDFKEDVIE 4255  
491 -----490  
4256 TIKYDGTDRRTAKEAMNPYSLDIFEDQLYWISKGEVWKNKFGQKKEKTLVVPWL 4315  
491 -----CL-----Y-----T-----494  
4316 TVRIFPHQLRYNKSVPNLCKQICSHLCLLRPGGYSACPGSSFTIEGSTTECDAATLPI 4375  
495 -LP-----H-----497  
4376 NLPPCRCHMGNCYFDTDLPKCKPGSGYTGKCEMAFSKGISPGTTAVALLTILLIV 4435  
498 -----H-----LV-----V-----I-----502  
4436 VIGALAIAGFFHVRRTGSLPLPALPKPLSLSSLVKPSSENGVTFRSGADLNMDIGVSGFG 4495  
503 P-----LVA-----LI-E--LL-----VLK 514  
4496 PETADRSAMSEDFVWEMGKQPIIPENPMYSARDSAV-K 4534

## RESULT 6

US-08-652-877-88

: Sequence 88, Application US/08652877

: Patent No. 6187548

: GENERAL INFORMATION:

: APPLICANT: Akerstrom, Goran

: APPLICANT: Juhlin, Claes

: APPLICANT: Raak, Lars

: APPLICANT: Crumley, Gregg R.

: APPLICANT: Morse, Clarence C.

: APPLICANT: Murray, Edward M.

: APPLICANT: Hjalms, Goran

: TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments

: INVENTION: Thereof and DNA Encoding Same

: NUMBER OF SEQUENCES: 106

: CORRESPONDENCE ADDRESS:

: ADDRESSEE: Rhone-Poulenc Rorer Inc.

: STREET: 500 Arcola Rd., 3C43

: CITY: Collegeville

: STATE: PA

: COUNTRY: USA

: ZIP: 19426-0107

: COMPUTER READABLE FORM:

: MEDIUM TYPE: Floppy disk

: COMPUTER: Macintosh





QY 220 -----Q----- 220  
Db 1680 MYNIQPLGIVAVHPKQPNVPCAFSRCSHCLLSSQGPFPYSCVCPSGWSLSPDLLN 1739  
QY 221 -----D-----V-----P-----SN-----K-----DV----- 228  
Db 1740 CLRDDQPLITVQHIIIFIGISLNPVKSNDAWPIAGIQNGLDVEFDDAEQIYIYWNPG 1799  
QY 229 -----T-----S-----LDW-----N-----S-----E-----G-----T 238  
Db 1800 EIHVKVTGNTRTVFASISVMGFSMMLALDWISRLYSTNPRQTSIEVLTLHGDIRYKT 1859  
QY 239 LLA-----T-----G-----S-----Y-----DG-----F----- 248  
Db 1860 LIANDGTALGVGPFIGITVDPARGKLYWSDQGTSDGVPKAIASANDGTSVKTFLTCNLE 1919  
QY 249 -----A-----R-----I-----W-----T----- 253  
Db 1920 HLECVTLIDIEQKLYWAVTGRGVIERNVDGTDRLMILVHQLSHPWGLIAVHDSFLYYTDEQ 1979  
QY 254 -----K-----D-----G-----N-----L-----ASTL-----G-----Q-----H-----K----- 266  
Db 1980 YEVIERVDKATGANKIVLRDNVFNRLGLQVYHRRNAAESSNGCSNNMACQQICLPVPG 2039  
QY 267 -----G-----P-----I-----FALK----- 273  
Db 2040 LFCACATGPKLPDNRSCSPYNSFIVVMSLAIKRGFSLESDHSETMVPVAGQGRNALH 2099  
QY 274 -----W-----N-----K-----GN----- 279  
Db 2100 VDVSVSGFIWYCDFSSVASDANAIKPKDGSLLMIVTHGIGENGVRGIAVDWAGNL 2159  
QY 280 -----I-----L-----S-----A-----G----- 285  
Db 2160 YFTNAFVSETLIEVIRINTTYYRLLKVTVMPRHVVDPKNRYLFWADYQRPKIERF 2219  
QY 281 -----I-----L-----S-----A-----G----- 285  
Db 2220 LDCNRTVLSGIVTPRGIAVDRSDGYVWVDDSLDIARIINGENSEVIRYGSRYPT 2279  
QY 286 -----VD-----K-----T-----TII-----W-----D----- 294  
Db 2280 PYGITVFENSIWDRNLKXIFQASKEPENTPEPTVIRDNINLWRTVIEDKQVQPRSPA 2339  
QY 295 -----K-----A-----HT-----G-----E-----A----- 300  
Db 2340 EVNNPCLNNGGCSHLCPALPLHTPKDCAFGLQSDGKNCAISTENFLIFALSNSLR 2399  
QY 301 -----K-----Q-----Q-----F-----P----- 305  
Db 2400 SLHLDPNHSPFQTIINVERTVMSLDYSDSDRIYFTQNLASGVQISYATLSSGIHTPT 2459  
QY 306 -----F----- 306  
Db 2460 VIASGIGTADGIAFDWITRIIYSDYLNQMINMAEDGSNRTVIARVPKPRAIVLDPQ 2519  
QY 307 -----H-----S-----AP-----ALD----- 313  
Db 2520 YLYWADWDTHAKIERATLGGNFRVPIVNSLSVMSGLTLDYEEDLLYWDASLQRIERST 2579  
QY 314 -----VD-----W-----Q-----S-----N----- 319  
Db 2580 LTGVDRVINVAHAFGLTLYGOYIYWDLTQYRIYRANKYDGGQIAMTLLSQPRG 2639  
QY 320 -----NT-----F----- 322  
Db 2640 INTVVNQKQCCNPNCEQNGGCSHICAPGNGAECQCPHEGNYLANNRKHCIVDNGER 2699  
QY 323 -----AS-----CS-----TD-----M-----C-----IH-----V----- 333  
Db 2700 CGASSFTCSNRCIIEEWKDNNDGCGDSEMSVCAHTCSPTATCANGRCVQYSYR 2759  
QY 334 CKL-----G-----QD-----R-----P-----I----- 342

Db 2760 CDYYNDGDSGDEACGLFRDCNATTEFCNNRRRCIPREFICNGVDNCHDNNNTSDEKNCPD 2819  
QY 343 -----K-----T-----FQ----- 346  
Db 2820 RTCQSGYTKCHNSNICIPRVYLCGDNDCGNSDENPTYCTHTTCSSEFQACSGRCIPQ 2879  
QY 347 -----GH-----T----- 349  
Db 2880 HWYCDQETDCFPASDEPASCSEHSEPTCLADEPKDGGRCIPSEWICDGDNDGDMSEDK 2939  
QY 350 -----N-----E-----VNA-----IK-----W----- 357  
Db 2940 RHQCQNQNSDSEFLCVNDPRDRRCIPQSMVWCDGVDCTDGYDENQNCNTRTCSNEFT 2999  
QY 358 ----- 357  
Db 3000 CGYGLCIPKIPRCDRHNDGYSDBRGCLYQTCQOQNTCQNGRCISKTFVCDENDCGD 3059  
QY 358 -----D-----PT-----G-----N-----L----- 363  
Db 3060 GSDELMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCNHLDDCLDSDKGCINECHDPSI 3119  
QY 364 -----LAS-----CS-----D-----D-----M----- 371  
Db 3120 SGCDHNCTDLTSTFYCSCRPGYKLSMDKRTCDVIDECTEMPFVCSQKCNVIGSYICKA 3179  
QY 372 ----- 371  
Db 3180 PGYLREPDKTCRONSINIEPYLIFSNRYLRNLITIDGYFYSILLEGLDNVVALDFRVEK 3239  
QY 372 -----T-----L-----K-----I-----W-----S----- 377  
Db 3240 RLYWIDTQOVIEMFLNKTNETIINHRLPAESLAVDWVSRKLYWLARDLGLFVSDL 3299  
QY 378 -----M-----KQ-----D-----N-----C-----VHDLQO-----NK----- 392  
Db 3300 NGHRRMLAQHCVDANNTFCFDPNPRGLAH--POYGYLYWADWGHRAHYGRVGMDCNTKS 3357  
QY 393 -----E-----I-----YT-----IK----- 398  
Db 3358 VIISTKLEWPNGITIDYNDLLYWADAHLGYEYSDLEGHHRHTVYDGLPHFPFAITIE 3417  
QY 399 -----W-----S-----P-----T-----G----- 403  
Db 3418 DIIYWDNTRIVKEGKNGYDGSNRQTLVNTTHRPFDIHVYHPYRQPIVSNPCGTNNGGCS 3477  
QY 404 -----PG-----T-----NN----- 408  
Db 3478 HLCLIXPGKGFTCECPDPTLQLSGSTYCMPCSSOTFLCANNEKCIPIWKKDQKD 3537  
QY 409 -----P-----NA-----N-----LML-----AS 417  
Db 3538 CSDGSELALCPORFCRLGQFCQSGNCTSPQTLNQHONCPDGSDEDL--LCENHHCDS 3596  
QY 418 -----A-----S-----F-----DS-----T-----VR-----L-----W-----DV 429  
Db 3597 NEWQANKRCIPESWQCDTFNDCEDNSEDSSSHCASRTCRPGQFRCANGRCIPQAWKCDV 3656  
QY 430 -----DRG-----I-----CI-----H-----T-----L----- 438  
Db 3657 DNDGDSDEPTEBECSSAHLCDNFTEFSCKTNYRCIPKAVCNVDDCDKDNDEQCEE 3716  
QY 439 -----T-----K-----H-----Q-----E-----P-----V----- 445  
Db 3717 RTCHPVGDPECKNHHCIPLRWQCDGQNDGDNSEENCAPRECTESEFRVCVQCIPSRW 3776  
QY 446 -----Y-----S-----V-----A-----F 450  
Db 3777 ICDHYNDGNSDERDCEMRTCHPEYFQCTSGHCHVSELKCDGSADCLDASDADCPTRF 3836  
QY 451 SPGRYL--AS-----GS-----F----- 462

Db 3837 -PDGAYQATMFECKNHVCIPYWKCDGDDCGSDGDEELHCLDVPKNSPNRFRCDNNR 3895  
QY 463 -D-----K-----CV-H-I-----468  
Db 3896 CIYSHEVCNGVDCGDTDEBEHCKRTPKPKCTEYKCGNGHCIPHDNVCDADDDCGD 3955  
QY 469 W-----N-TQ-----V-----CL-----475  
Db 3956 WSDLGCKGKERTCAENICQNTQNEGFGICCTAGFTNVFRTSCLDINECEQFG 4015  
QY 476 -H-----476  
Db 4016 TCPQHCRNTKGSYECVADGFTSMRDPKRCACAEGLSPDLLPDNVRIRKYNLSRERS 4075  
QY 477 -YL-----N-G-----480  
Db 4076 EYLQDEYIQAVDYDWDPEDIGLSVYVYVVRGSGRFGAIKRAYINFESGRNLLQVEVD 4135  
QY 481 -Q-----V-----L-----N-----485  
Db 4136 LKLKYYMQDGIADVWVGRHRYWSDVKNKRIEVAKLDGRYRWLISDTLDQPAIAVNP 4195  
QY 486 LG-----R-----SI-----490  
Db 4196 LGLMFTWDMGKPKXESAMNGEDRNILVFEDLGWPTGLSIDYLNNDRIYWSDFKEDVIE 4255  
QY 491 -----490  
Db 4256 TIKYDGTDRRVIAKEAWNPYSLDIFEDOLYISKEGEVWKQKFGGKKEKTLVNPWL 4315  
QY 491 -----CL-----Y-----T-----494  
Db 4316 TQVRIFQLRYNKSVPNLCKQICHLCLLRPGYSACPCQGSFIEGSTECDAAIPLI 4375  
QY 495 -LP-----H-----497  
Db 4376 NUPPPRCMHGNCYPDETDLPKCKPFGYTGKCEMAFSKGISPGTTAVALLTILLIV 4435  
QY 498 -----H-----LV-----V-----I-----502  
Db 4436 VICALAIAGFFHYRRTGSLLPALPKLPSSLVLPSENGVTFPSGADLNMDIGVSGFG 4495  
QY 503 P-----LVA-----LI-E--LL-----VLK 514  
Db 4496 PETADRSMAMSEDFVMEKGQPIIPENPMYSARSAV-K 4534

RESULT 7

US-08-652-877-90  
; Sequence 90, Application US/08652877  
; Patent No. 6187548  
; GENERAL INFORMATION:  
; APPLICANT: Akerstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Raak, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjalms, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein; Fragments  
; TITLE OF INVENTION: Thereof and DNA Encoding Same  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd., 3C43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.5.1

SOFTWARE: Word 6.0 (Patentin)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,877  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/15203  
FILING DATE: 22-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,836  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,314  
FILING DATE: 07-JUNE-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Savitzky, Martin  
REGISTRATION NUMBER: 29,699  
REFERENCE/DOCKET NUMBER: A1355E-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-454-3816  
TELEFAX: 610-454-3808  
INFORMATION FOR SEQ ID NO: 90:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4655 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-652-877-90

Query Match 70.1%; Score 2558.8; DB 3; Length 4655;  
Best Local Similarity 9.5%; Pred. No. 5.3e-32;  
Matches 430; Conservative 60; Mismatches 18; Indels 4032; Gaps 371;  
QY 1 M-----S-----I-----S-----S-DE- 7  
Db 1 MDGPAVACTLLALVACLAPASGQECDSAHFRCSGHCIIPADWRCGTGKDCDDADEI 60  
QY 8 -VN-----L-----V-----12  
Db 61 GCNAVTCQGYKQSGEGOCIPSSWVCDQDQDDGSDERQDCSQSTGSHQITCSNGQC 120  
QY 13 -YR-----Y-----L-----Q-----E---S 19  
Db 121 IPSEYRCDHVRCPDGADENDCQYPTCEQLTCDNGACYNTSQKCDWKVDCRSDSDEINCT 180  
QY 20 -G-FS-----H-----S-A-F-T-----F---G-I-----30  
Db 181 EICLHNEFCGNGECIPRAYVCDHNDQDQSDGDEHACNYTCGYOFTCPSGRCIYQNV 240  
QY 31 -K-----S-----H-----IS-----Q---SNI 39  
Db 241 CDGEDDCKNGDEDCESGPHDVHKCSPREWSPESGRCISYKVCDCGILDCPREDEN- 299  
QY 40 N--G-----AL-----VP-----PAALI-----S-----I-I-- 53  
Db 300 NTSTGKCSMTLCSALNCOYQCHETPYGACFCPPGYIINHDSRTCVDFDQCIWIGICD 359  
QY 54 QK-----G-L--QY-----VE 61  
Db 360 QKCESRPRGRLCHCEBEGYILERGQYKANDSPGEASIIIFNSGRDLLIGDINGHSRFLVE 419  
QY 62 -----A-----E---VSI-----N 67  
Db 420 SONRGVAVGVAFHYHLQRFVFTDVTQNVKVPFVSDINGLNIQEVNLNVSVEIPENLAVDWN 479  
QY 68 -----E-----DG-----TL-----72  
Db 480 KIYLVETKVNRI DMVNLDSYRVTLITENLHGRGIAVDPTVGYLFFSDWESLGSBPKE 539  
QY 73 -F-DG--R-----P-----IE-----S 80  
Db 540 RAFMDGSRNKKDLVKTGLGWPAGVTLMDISKRVYVWDSRFDYIETVTYDGIQRTVTVHGS 599



Db 2760 CDYNDGCGSDEAGCLFRDCNATTFBMCNRRRCIPREFICNGVDNCHDNTSDEKXCPD 2819  
QY 343 -----K-----T-----FQ----- 346  
Db 2820 RTCQGYTKHNSNICIPRYVLCDGNDGCDNSDENPTYCTHTCSSSEFOCASGRICIPQ 2879  
QY 347 -----GH-----T----- 349  
Db 2880 HWYCDQETCFDASDEPASCGHSERTCLADEFKDGGRCIPSEWICDGDNDGMSDEDK 2939  
QY 350 -----N-----E-----VNA-----IK-----W----- 357  
Db 2940 RHOCQONQCSDFLCVNDRPPRRRCIPQSFOSWVCDGVDCTGDYDENQNCNTRTCSNEFT 2999  
QY 358 ----- 357  
Db 3000 CGVGLCIPKIFPCRDHRNDGYSDERGCLYQTCQOQFTQNGRCISKTFVCDENDDNGD 3059  
QY 358 -----D-----PT-----G-----N-----L----- 363  
Db 3060 GSDELMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCNHLDDCLNDSBKGGCIGINECHDPSI 3119  
QY 364 -----LAS-----CS-----D-----D-----M----- 371  
Db 3120 SGCDHNCTDTLTSFYCSRCRPGYKMSDKRKTCDVIDECTEMPFVCSQKCNVIGSYICKCA 3179  
QY 372 ----- 371  
Db 3180 PGYLREPDKCTQONSNIEPYLIFSNRYLRNLTIDGYFVSLILEGLDNNVALDPORVEK 3239  
QY 372 -----T-----L-----K-----I-----W-----S----- 377  
Db 3240 RLWIDTQOVIEMFLNKTNETIINHRLPAESLAVDWVSRKLYLWLDARLDGLFVSDL 3299  
QY 378 -----M-----KQ-----D-----N-----C-----VHDLOQ-----H-----NK----- 392  
Db 3300 NGHREMLAQHCVDANNTCFDNPRLALH--PQYGYLWADWGHRAIYGRVGMGTNKS 3357  
QY 393 -----E-----I-----YT-----IK----- 398  
Db 3358 VIISTKLEWNGITIDYNDLLYWAHAHLGYIEYSLEGHHRHTVVDGALPHFPALTIFE 3417  
QY 399 -----W----- 3403  
Db 3418 DTIYTDWNRTRVEKNGKYDGSNRQTLVNTHTPEPDHVVHPYRQPIVSNPCGTNNGCS 3477  
QY 404 -----PG----- 408  
Db 3478 HLCILKPGGKFTCECPDDFRTLQLSGTYCMPMCSSTOPLCANNEKCIPIWKKCDGQXD 3537  
QY 409 -----LML----- 417  
Db 3538 CSDGSDELALCPORFRLGQFCSDGNGCTSPTQLCNAHNCNCPDGSDEDRLLCENHHCDS 3596  
QY 418 -----A-----S-----F-----DS-----T-----VR-----L-----W-----DV----- 429  
Db 3597 NEWQCANKRCIPESWQCDTFNDCEDNSDESSHCASRTRCPGQFRANGRCIPQAKKCDV 3656  
QY 430 -----DRG-----I-----CI-----H-----T-----L----- 438  
Db 3657 DNDGDSDEPIBECMSAHLCDNFTEFSCKTNYRCIPKWAQVGVDDCDRDSDEQCEE 3716  
QY 439 -----T-----K-----H-----Q-----E-----P-----V----- 445  
Db 3717 RTCHPVDFRCNHHCIPLRMQCDGNDGNDSEENCAPRECTESEFCVNOQCIPSRW 3776  
QY 446 -----Y-----S-----V-----A-----F----- 450  
Db 3777 ICDHYNDGDSNDERCEMRTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCTPF 3836  
QY 451 SPDGRYL-AS----- 462  
Db 3837 -PDGAYCOATMECKNHVCIPPYWKCDGDDCDGSDDEELHCLDVPNCSPNFRFRCDNNR 3895

QY 463 -----D-----K-----CV-H-----I----- 468  
Db 3896 CIYSHEVCNVDGCGDTDETEHCRKPTPKPCTEYKCGNGHCIPHDNVCDADDCD 3955  
QY 469 W-----N-----TO-----V-----CL----- 475  
Db 3956 WSDLGCKNGKERTCAENICEONCTQNBGGFTCSCTAGFTNVFRTSCLDINECEQFG 4015  
QY 476 -----H----- 476  
Db 4016 TPOHCRNTKGYECVACADGFTSMRDRPKRCAAEGSSPLLLPDNVNRIRKYNLSRFRS 4075  
QY 477 -YL----- 480  
Db 4076 EYLQDEEYIQAVDYDWDPEDIGLSVVVYTVRGESRFGAIKRAYIPNFESGRNLLVQEV 4135  
QY 481 -----Q-----V-----L-----L-----N----- 485  
Db 4136 LKLKYYMQPDGIAVDWVGRHIYMSDVKNKRIEVAKLDRYRWLISDTLDQPAIAVNP 4195  
QY 486 LG-----R-----SI----- 490  
Db 4196 LGLMFTWDGKBPKELESAMNGEDRNILVFDLGPWPTGLSIDYLNNDRIYWSDFKEDVIE 4255  
QY 491 ----- 490  
Db 4256 TIKYDGTDRRTAKEAMNPYSLDIPEDQLYISKEKEVWKQKFGQKKEKTLVNPWL 4315  
QY 491 -----CL-----Y-----T----- 494  
Db 4316 TQVRIFHQLRYNKSVPNLCKQICSHLCLLRPGGYSCAPQSGSFIEGSTTECDAIELPI 4375  
QY 495 -LP-----H----- 497  
Db 4376 NLPPRCRWHGNCYFDETDLPKCKPCPSGYTKYCEMAFSKGISPGTTAVAVLLTILIV 4435  
QY 498 -----H-----LV-----V-----I----- 502  
Db 4436 VIGALAIAGFFHYRRTGSLPALPKPLPSLSLVKPSSENGVTFRSGADLNMDIGVSGFG 4495  
QY 503 P-----LVA-----LI-E-----LL-----VLK 514  
Db 4496 PETADRSWMSGEDFVWEMGKQPIIFENPMYSARDSAV-K 4534

## RESULT 8

US-08-652-877-86  
; Sequence 86, Application US/08652877  
; Patent No. 6187548

## GENERAL INFORMATION:

APPLICANT: Akerstrom, Goran  
APPLICANT: Juhlin, Claes  
APPLICANT: Raak, Lars  
APPLICANT: Crumley, Gregg R.  
APPLICANT: Morse, Clarence C.  
APPLICANT: Murray, Edward M.  
APPLICANT: Hjalms, Goran

TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
TITLE OF INVENTION: Thereof and DNA Encoding Same  
NUMBER OF SEQUENCES: 106  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Rd., 3C43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA

ZIP: 19426-0107

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: Macintosh

OPERATING SYSTEM: System 7.5.1

SOFTWARE: Word 6.0 (patentin)



Db 1680 MYNIQWPLGIVAVHPKQPNVNCAPAFRCRSHLCLSSQGHFYSCVPCPSGWSLSPDLLN 1739  
QY 221 ---D---V---P---SN---K---DV--- 228  
Db 1740 CLRDDQPLITVQHIIFIGISLNPVKSNDAMVPIAGIQNGLDVEFDDABQYIYWENPG 1799  
QY 229 ---T---S---LDW---N---S---B---G---T 238  
Db 1800 EIHVRKTDGTRTFVAFASIMVGPMSMLALDWISRNLYSTNPTQOSIEVLTLHGDIRYKRT 1859  
QY 239 LLA---T---G---S---Y---DG---F--- 248  
Db 1860 LIANDGTALGVGPPIGITVDPARGKLYWSDQDTSVGPAPAKIASANMDGTSVKTLFTGNLE 1919  
QY 249 ---A---R---I---W---T--- 253  
Db 1920 HLECVTLIDEEQKLYWAVTCRGVIERGVNDGTDMILVHQLSHPWGLIAVHDSFLYTDEQ 1979  
QY 254 ---K---D---G---N---L---ASTL---G---Q---H---K--- 266  
Db 1980 YEVIERVDRATGANKIVLRDNPVNLRLQVYHRRNAEASSNGCSNNMNAQQICLPVPGG 2039  
QY 267 ---G---P---I---FALK--- 273  
Db 2040 LFSCACATGKLPNDRNSCPYNSFIWVMSLAIKRGFSLESDHSETMVPVAGQGRNALH 2099  
QY 274 ---W---N---K---GN--- 279  
Db 2100 VDVVSSGFIYWCDFSSVASDANAIKPKDGSLLMNIIVTHGIGENGVRGIAVDWVAGNL 2159  
QY 280 ---I---L---S---A---GN---F 280  
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTVDMPRHIVDPKRYLFWADYQRPKIERSF 2219  
QY 281 ---I---L---S---A---G--- 285  
Db 2220 LDCTNRTVLVSEGITVPRGLAVDRSDGYVYVDDSLDIIRIRINGENSEVIRYGRYPT 2279  
QY 286 ---VD---K---T---TII---W---D--- 294  
Db 2280 PYGITVPENSIWDRNLKIKIPOASKEPENTPEPTVIRDNINMLRDTVFDKQVQRPSPA 2339  
QY 295 ---A---HT---G---E---A--- 300  
Db 2340 EVNNPCLNNGCCHLCPALPGLHPCPKDCAFGLTQSDGKNCAISTENFLIFALSNSLR 2399  
QY 301 ---K---Q---Q---F---P--- 305  
Db 2400 SLHLDPENHSPPTQINVERTVMSLDYSDVSDRIYFTONLASGVQISYATLSSGIHTPT 2459  
QY 306 ---F--- 306  
Db 2460 VIASGIGTAGIAFOWITRIIYSDYLNQMINMAEDGNRTVIARVPKRAIIVLDPQCG 2519  
QY 307 ---H---S---AP---ALD--- 313  
Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSLVMPSPGLTLDYBEDLLYWDASLQRIERT 2579  
QY 314 ---VD---W---Q---S---N--- 319  
Db 2580 LTGVDRREVINAHAVAFGLTLYQYIYWTDLTYQRIYRANKYDGSQGIAMTTLNLSQPRG 2639  
QY 320 ---NT---F--- 322  
Db 2640 INTVKNQKQCNPCQENGGCSHICAPGNGAECQCPEGWYLIANNRKHCIVDNGER 2699  
QY 323 ---AS---CS---TD---M---C---IH---V--- 333  
Db 2700 CGASSFTCSNGRCISEWKDNDNDGCGSDMESVCAHTCSPTAFTCANGRCVQYSYR 2759  
QY 334 CKL---G---OD---R---P---I--- 342  
Db 2760 CDYNDGCGSDEAGLFRDCNATTEFMCNRRRCIPREFICNGVDNCHDNTSDEKNCPD 2819

QY 343 ---K---T---FQ--- 346  
Db 2820 RTCQSGYTKCHNSNICIPRYVLCDGNDGCDNSDENPTYCTTHTCSSEFQCTSGRCIPIQ 2879  
QY 347 ---GH---T--- 349  
Db 2880 HMYCQDETDFDASDASPACGHSERCTLADEFKCDGRCIPSEWICDGDNDGDMDSDEK 2939  
QY 350 ---N---E---VNA---IK---W--- 357  
Db 2940 RHQONQNSDSEFLCNDRPPDRRCIPQSWVWCDGVDCTDGYDENQNCNTRTTCSENEFT 2999  
QY 358 --- 357  
Db 3000 CGYGLCIPKIFRCDRHNDGCDYSDERGCLYQTCQOQNFQCONGRCISKTFVCDENDDCGD 3059  
QY 358 ---D---PT---G---N---L--- 363  
Db 3060 GSDELMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCNHLDDCLDNDSEKGGCINECHDPSI 3119  
QY 364 ---LAS---CS---D---D---M--- 371  
Db 3120 SGCDHNCTDLTSFYSCRPYKLMDSKRTCVDIDECTEMPFVCSQKCNENVISYICKCA 3179  
QY 372 --- 371  
Db 3180 PGYLRBPDKTCRQNSNIEPYLIFSNRYVLRNLTDIGYFSLILEGLDNNVALDFORVEK 3239  
QY 372 ---T---L---K---I---W---S--- 377  
Db 3240 RLYWIDTQVIERFMFNKTKETIINHRLPAAESLAVDWVSRKLYWLDAARLDGLFVSDL 3299  
QY 378 ---M---KQ---D---N---C---VHDLOQ---H---NK--- 392  
Db 3300 NGGHRMLAQHCVDANNTFCFDPNPRGLAH---PQGYLYWADMGHRAIYIGRVGMDGTNS 3357  
QY 393 ---E---I---YT---IK--- 398  
Db 3358 VLIISTKLEWPNGITDITDNDLLYWADAHILGYIEYSDLEGHRRHTVYDGLALPHFAITIFE 3417  
QY 399 ---W---S---P---T---G--- 403  
Db 3418 DTIYTDWNTRTVEKNGKYDGSNRQTLVNTTHRPFDIHVYHPYRQPIVSNPCGTNNGCS 3477  
QY 404 ---PG---T---NN--- 408  
Db 3478 HLCLIKPGKGFTCECPDPTLQLSGSTYCMPCSSSTQFLCANNEKCIPIWKKCDGQKD 3537  
QY 409 ---P---NA---N---LML---AS 417  
Db 3538 CSDGDELALCPQRCRLQFQCGSDGNCSTSPOTLCNAHQNCPPDGSDEDL---LCENHHCS 3596  
QY 418 ---A---S---F---DS---T---VR---L---W---DV 429  
Db 3597 NEWQCANKECIPESWOCDFNDCEDESDESHCASRTCRPGCFRCANGRCIPQANKCDV 3656  
QY 430 ---DRG---I---CI---H---T---L--- 438  
Db 3657 DNDGDSHDEPIEECMSSAHLCDNFTEFSCKTNYRCIPKWAVCNGVDDCRDNDSEQGBE 3716  
QY 439 ---T---K---H---Q---E---P---V--- 445  
Db 3717 RTCHPVGDPRCNHHCIPURWOCQDNDGCDNSDENCAPRECTSEFRVCNQCIPSRW 3776  
QY 446 ---Y---S---V---A---F 450  
Db 3777 ICDHYNDGCDNSDERDCERMTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCTPRF 3836  
QY 451 SPDGYL---AS---GS---F--- 462  
Db 3837 -PDGAYCOATMECKNHVICIPPYWKCDGDDDCGSDGDEBELHCLDVPNCSPNFRCDNNR 3895

```
QY 463 -----D-----K-----CV-H-I----- 468
Db 3896 CIYSHEVNGVDCGDTDETEHCKRTPKPCTEYKCGNGHCIPHDNVCDADDGCD 3955
QY 469 W-----N-TQ-----V-----CL----- 475
Db 3956 WDELGNKGRKERTCAENICEQNTQLNEGGFICSTAGFETVNPDRTSCLDINECEQFG 4015
QY 476 -----H----- 476
Db 4016 TCPQHCNRTKGYECVACDGTSMRPGKRCACAGSSPLLLPDNVRIRKYNLSRERS 4075
QY 477 -YL-----N-----G----- 480
Db 4076 EYLQDEYIOAVDWDPEXDIGLVVYTVRSGSRFGAIKRAYIENFESGRNLVQEV 4135
QY 481 -----Q-----V-----L-----N----- 485
Db 4136 LKLKYMOPDGIADVGRHIYMSDVKNKRIEVAKLDGRYRWLSTDLDPAAIAVNP 4195
QY 486 LG-----R-----SI----- 490
Db 4196 LGLMFWDMGKPKXBSAMNGEDRNILVPEDLGPWPTGLSIDYLNNDRIYMSDFKEDVIE 4255
QY 491 ----- 490
Db 4256 TIKYDGTDRRVIKAEAMNPSYLDIFEDQYIWSKEGEVWKQNGKGGKKEKTLVNPWL 4315
QY 491 -----CL-----Y-----T----- 494
Db 4316 TVRIFHQLRYKSVNPLCKQICSHLCLLRPGGYSCACPOGSGFIEGTTCECDAAIELPI 4375
QY 495 -LP-----H----- 497
Db 4376 NLPPPCRMHGNCYFDETLDPKCKPSGYTKYCEMAFSGISPGTTAVAVLLTILLIV 4435
QY 498 -----H-----LV-----V-----I----- 502
Db 4436 VIGALATAGFYHRTGSLPALPKPLSLKVPSENGVTFRSGADLNMDIGVSGFG 4495
QY 503 P-----LVA-----LI-E--LL-----VLK 514
Db 4496 PETAIIDRSMAMSEDFVMEMGKQPIIFENPMYSARDSAV-K 4534

RESULT 9
US-09-334-220-1
; Sequence 1, Application US/09334220
; Patent No. 6323177
; GENERAL INFORMATION:
; APPLICANT: St. Jude's Children's Research Hospital
; APPLICANT: Curran, Thomas
; APPLICANT: D'Arcangelo, Gabriella
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND
; TITLE OF INVENTION: THERAPIES
; FILE REFERENCE: 2427/0F704
; CURRENT APPLICATION NUMBER: US/09/334,220
; CURRENT FILING DATE: 1999-06-16
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 3460
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-334-220-1

Query Match 70.0%; Score 2554; DB 3; Length 3460;
Best Local Similarity 11.3%; Pred. No. 1.3e-32;
Matches 391; Conservative 95; Mismatches 14; Indels 2950; Gaps 357;

QY 1 M--S-----I-----S-----S-----D-----EV----- 8
```

```
Db 1 MERSGARQTFLLALLGATLRARAAAGYPRFSPFFFLCTHHGLEGDGGQGEVLISLH 60
QY 9 --N-----F--LV--VRY--LQ--ES-----GF--S-H- 23
Db 61 IAGNPTVYVPGQEVHTVISTSTFPDGLLVGL-YTSTVQASOSIGGSAFGFGIMSDHQ 119
QY 24 -----S-A-----F-----T-----F-----G-I-K----- 31
Db 120 FGNQFMCVWASHVSHLPTTNLSFIWAPPAGTCGVNFMATATHRGQVIFKDALAQOLCE 179
QY 32 -----S-----H-----I-----S-----Q--SNI--N-----GAL----- 43
Db 180 QGAPTDVTVHPLABIHSDSIILRDDPDSYHQQLNPNINWECNNCETGEOCGAIMHGNA 239
QY 44 V-----P--P-----A-AL--I-----SI-----IQ----- 54
Db 240 VTFCEPYGRELITGLNLTASVLOFSGSCRSFSDPSIIIVLYAKNSADWIOLEK 299
QY 55 -----KG--LQY--V-EA-E-----VSIN----- 67
Db 300 IRAPSNVSTIIHILYLPEDAKGENVQFQWQENLRVGEVYACWALDNILI-INSARQV 358
QY 68 --ED-----GT-IF-----DG-----R----- 76
Db 359 VLESLDPVDTGNWLFPPGATVKHSCQSDGNSIYFHGEGSEFNFATTRDVLSTEDIOE 418
QY 77 -----P-----IES--LSLI-----DA----- 86
Db 419 QWSEFESOPTGWDVLGAVIGTECGTIESGLSVFLKDGKELCTPMSDITGYGNLRFYF 478
QY 87 VM-----P-----D-----D-----V-----QT----- 94
Db 479 VMGGICDPCNSHENDIILYAKIEGRKEHITLDTLSYSSYKVPKPSLVSVINPELOTPATFK 538
QY 95 --RQ--QAY-R-----D-----K-LAQ-----Q----- 106
Db 539 CLRQKHQGNRNVWAVDFHVLVPLPSTMHMIQPSINLGCCTHQPQGNVSLEFSTNHG 598
QY 107 QA-----A-----AAA-----AA----- 114
Db 599 RSWSLHTTECLPEICAGPLPHSTVSSYSENYSNWNRIITPLNAAALTRNTRIRWRGTPI 658
QY 115 -----A-----A-AASQ-----Q--GSA 124
Db 659 LGNMWAIENVYIGPSCLKFCGRGQCTRHGCKDCDFGSPACEMASQTPPMFISEFGSS 718
QY 125 K-----N-----G-E-----N----- 129
Db 719 RLSSYHNFYSIRGAESVFGCVLASGKALVFNKEGRRLITSPDSSQSRFLOFTLRGSG 778
QY 130 -----T--A-----NG-----E-----E--N--G-A----- 138
Db 779 KSVLSTCRAPDQPGEGVLLHYSYDNGITWKLLEHYSYLSVHEPRIIISVELPGDAKQFGIQ 838
QY 139 -----H-----T-----I-----AN-----N-----H-- 145
Db 839 FRWWQPYHQSOREDVWAIIDBIIMTSLVFNISLDFTLNVEVTQSLGYLGNVOPYCCHDW 898
QY 146 -----T-D-----M-----MEVDG-----D-V-E----- 156
Db 899 TLCFTGDSKLASSNRVVFETQSMQI-GASYMIQPSLVMGCGQKYTPHMDNQVKLEYSTNHG 957
QY 157 -----I-----P--PN--K-----A-----VVL-----R----- 166
Db 958 LTHVLVQEELCPMSPSCQEFTSASIVHASEFTQWRRVIVLLPQKTWSSATFRFWSOYYT 1017
QY 167 -----CH-----E-----SE-----V--F-- 173
Db 1018 AQDEWALDSIYIQOQCPNMCSGHGDHICRCDDQYQGTECHPEALPSTIMSDFNQN 1077
QY 174 -----I-----C-----A-----W----- 177
Db 1078 GWESDWQEVIGGEIVKPEQCGVSSGSSLYFSKAGKRLQVMDLDTSWDFVQFYIQIG 1137
```



QY 178 ---N-PVSD---LIA-S--GSG---D-S---T-- 192  
Db 1138 GESASCNKP--DSREBGLVLOYNNG-GIOWHLLAEMYFSDFKPRFVYLELPAAKTPC 1194  
QY 193 AR-----I-----W-----N-L-----S-199  
Db 1195 TRFRWQPVFSGEDYQWAVDDIIILSEKQKQIIPVINPTLPQNFYKPAFYPMQMSV 1254  
QY 200 ---E---N---ST-S--G-S---T---Q-L-----209  
Db 1255 WMLANEGWVNETFCAATPSAMIFKSGDGRFAVTRDLTKPGYVLQFKLNTGCANQPS 1314  
QY 210 ---V-LR-H---CIREG---GOD-----221  
Db 1315 STAPVLLOYSHDAGMSWFLVKEGYPASACKGC--EGNSRELSEPTMYHTG-DFEBWTRI 1371  
QY 222 ---VP-----S-N-----K-----D---V-----228  
Db 1372 TIVIPRSLASSKTRFRWIOBSSSQKNVPPGLDGVYISEPCPSYSGHGDICISGVCFCDL 1431  
QY 229 ---T-S---LD-----W-----N---S-----EG-- 237  
Db 1432 GYTAAQCTCVSNPNENPFREGKSLPYKITGAQVGTGCTLNDGKSLYFNGPKR 1491  
QY 238 ---TL-LAT---GS-----Y-D-G-----247  
Db 1492 EARTVPLDTNIRLVQFYIQGSKTSGITCKPRTRNEGLIVQYSDNGILWHLLBELDF 1551  
QY 248 ---F-----A-RW-----T---K-- 254  
Db 1552 MSPLPQIISIDLPODAKTPATAFR-WWOPQHKGSAQWALDDVLIGMNDSSQTGFQDKF 1610  
QY 255 DG-----LAS-T-L---G-Q-----H-----265  
Db 1611 DGSIDLOANWYRIQGGQVDDICL-SMDTALIFTENIGKPRYAETWDFHVSASTFLOFEMS 1669  
QY 266 KG---PIF-----AL--K-W--NKK--G-----N-----279  
Db 1670 MGCKSP-FNSHVSQVLOYSLNKGDKHLVTEBCVPPTIGLHYTESIYTSERFQWKR 1728  
QY 280 ---F-I--LSAGVDKT-TI--I-----W-----DA-----295  
Db 1729 TVYLPLSTISPRFRWIOQANYTVGAD-SWAIDNVVLASGCPWMSGRCIGDAGRCVDR 1787  
QY 296 ---H-----T-----CE--- 299  
Db 1788 GFGGPYCVVPVPLPSILKDDFNGLHPLDLWPEVYGAERGNLNGETIKSGTSLIFKGEGLR 1847  
QY 300 ---AK-Q--QF---P-----F-----H-----SA--- 309  
Db 1848 MLISRDLCTNTMYVQPSLRPIAKSPERSHILLOFSISGGITWHLMDIFYPPQTINIL 1907  
QY 310 ---P-----A-L-----D-----V-----314  
Db 1908 FINVPLPYTAQTNATFRMLWQPNYNGKKEIWIIVDDFIIDGNVNVNPNVLLDFFDGPPE 1967  
QY 315 D-W-----Q-----SN-----NT--P--A 323  
Db 1968 DNWFYPPGNIGLYCPSYKGAPEDSAMVFSVNEVGEHSITTRDLNVNENIIQFENV 2027  
QY 324 SCSTD-----MC-----330  
Db 2028 GCSSTDSSADPVLEFRDFGATWHLPLCYHSSSHVSLCSTEHPSTYYAGTNQGW 2087  
QY 331 ---IH-----V-----333  
Db 2088 RREVWHFGLKHLGSRVFRWYQGYFAGSQPVTAIDNVYVIGQCEMCMNGQSGCINGTK 2147  
QY 334 ---CK-----L-----GO--DR-----P-----I-KT-----P-Q 346  
Db 2148 CIGDPYSGPTCKISTKNPDLKDDFEGQLESDFLLMSGKPKSRCKGILSSGNLFFNE 2207

QY 347 -G-----H-----T-N-----EV--- 352  
Db 2208 DGLRMLTRDLDLSHARFVQFFMRLCGKGVPDRSQPVLLQYSLNGLSLSLLQELPFS 2267  
QY 353 ---N---A-I--K-----W-----357  
Db 2268 NSSNVGRYIALEIPLKARSGSTRLRWQPSENGHFYSPWVIDQILIGNISGNTVLBDDF 2327  
QY 358 ---D-----P---TG-----N--L--ASCS 368  
Db 2328 TTLDSRKWLLHPGQTKMPVCGSTGDALVFIEKASTRYVYSTDVAVNEDSLQIDFAASCS 2387  
QY 369 ---D-----DM-----T-----L-KI-----W-----376  
Db 2388 VTDSCVAIELEYSDVLGLSWHPLVRDCLPTNVBCSRVHLQRIILVSDTFNKMTRITLPLPP 2447  
QY 377 ---S---M-----KQ---DN---CV---H-----385  
Db 2448 YTRSQATRFRWHPQAPFDKQOTWAIIDNVYIGDCIDMCSGHGRICQNCVCDEQWGLYC 2507  
QY 386 -D---L-Q---Q-----Q-----H-----390  
Db 2508 DDPETSLPTQLKDNFNRAPSSQNLTVNGKSLTVCGAVASGMALHPSGCGSRLLVTVDL 2567  
QY 391 ---NKE-I---Y---T-----IKW-----S-PTG--- 403  
Db 2568 NLTAEBFIQPFMYGCLITPNRNOGVLLLEYSVNGGITWLLMEIFVDOYKSKP-GFVNIL 2626  
QY 404 ---P-G-TN---N-----ML-----A--- 416  
Db 2627 LPPDAKEIATRFRWQPRHDGLDQNDWAIDNVLIGSADQRTVMLDTFSSAPVQHERSP 2686  
QY 410 -NA-----N--L-----ML-----A--- 416  
Db 2687 ADAGPVGRIAFDMFMDKTSVNEHMLFHDCTVERFCDSPDGVMLCGSHDREVIYVTHD 2746  
QY 417 ---S---A-----S--F-----D-----421  
Db 2747 LPTEGMIQFKISVCKVSEKIAQNIHQVYSTDFGVSNVYLPOCLPADPKCSGVSQ 2806  
QY 422 -S-----T-----VR---L---W-D-----VD-RG--- 432  
Db 2807 PSVFFPTKGKRITYPLPSLVGNVFRFYQKYSMDQMAIDNFYLGPGCLDNCRGGDC 2866  
QY 433 ---IC-----I-HL-T-KHQ---E-P-----V-----445  
Db 2867 LREQICDPYSGPNCYLTHLTKFLKRFDPSEEIKPDLWMSLEGGSTCTECGILAEFTA 2926  
QY 446 -Y--S-----V-----A-F-----S-P-----DG--- 454  
Db 2927 LYFGGSTVQAVTQDLDLRGAKFLQYWGRIGSNNMTSCHRPICRKEGVLLDYSTDGGIT 2986  
QY 455 ---R-Y--L--A-----SG-----460  
Db 2987 WTLHEMDYQKYSVVRHDYILLPEDALTNTTLRWQPFVISINGIVVSGVERAQWALDNI 3046  
QY 461 ---S---PD-----K-----C---VHI-W-N--T--- 471  
Db 3047 LIGGAEINPSQLVDTDEGTSGHEENWSPYNAVTRTAGFCGNPSFHYLPNKKKDKTHNA 3106  
QY 472 ---Q---V---C-LH---Y-----L---N---479  
Db 3107 LSSRELI IQGYMMQPKIVVGCEATSCGDLHSWMLEYTKDARSQNLVQTCPLPSSNS 3166  
QY 480 -G---Q---VL--LN-----485  
Db 3167 IGCSPQFHEATYNSVNSSSWKIRITIQLPDHVSSSATQFRWIKGEBETEKQSWAIDHVV 3226  
QY 486 ---L-----G-----R-----SI 490  
Db 3227 IGEAPKLCSGHGYCTTGTAICIDCESFQGDGCVSFHSHDLPSYIKONFESARVTEANWETI 3286  
QY 491 ---C-----LY-----T-----494

Db 3287 QGGVIGSCQQLAPYAHGDSIFYNGCQIROATKPLDLTRASKIMFVLQIGSMQSDTSCN 3346  
QY 495 --L--PH-----HLV-----V-----IPL-----V----- 505  
Db 3347 SDSLGGPDAVDKAVLLQYVNNGITWHVIAHQPKDFTQAKRQSVNVVPLEARMKGVLLRW 3406  
QY 506 -----AL--IE--LLV-----LK 514  
Db 3407 QPRHNGTGHODWALDHVEVVLVSTRKQNYM 3436  
RESULT 10  
US-08-718-388-9  
; Sequence 9, Application US/09718388  
; Patent No. 6271362  
; GENERAL INFORMATION:  
; APPLICANT: MORIKAWA, MINORU  
; APPLICANT: HARADA, NAKKI  
; TITLE OF INVENTION: GENE ENCODING IGG Fc REGION-BINDING  
; TITLE OF INVENTION: PROTEIN  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/718,388  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 0230-111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5405 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-718-388-9

Query Match 69.9%; Score 2552.2; DB 3; Length 5405;  
Best Local Similarity 9.7%; Pred. No. 1.6e-31;  
Matches 408; Conservative  
QY 1 M-----S-I-----SS-D-EVN-----FLV-----YR-----Y----- 15  
Db 1 MGALWSWILWAGATLLWGLTQESVDLK-NTGREBFLTAFLQNYQLAYSKEYPRLLISS 59  
QY 16 -----L-O-----ES-----G--F--S--H----- 23  
Db 60 LSESPASVLSQADNTSKVTVRPGESVMWNISAKAEMIGSKIFOHAVVIHSDYALSQ 119  
QY 24 -----SA-----F-----T 27  
Db 120 ALNAKPDABELTLRPIQALGTEVFLVTPPGTSARNVKEFAVVAGAAGASVVTLKGSVT 179  
QY 28 F-----G-----I-----K-----S-----HI- 34  
Db 180 FNGKFYPAGDVLRTLPQYNNVAQLQSSVDLSGSKVTASSPFAVLSGHSCAQKHTTCNHV 239

QY 35 -----SQS-----N--I----- 39  
Db 240 EQLLPTSAMGTHVVTPLASQSRDYDLAFVVASQATKLTYNHGGITSGRGLQAGDVVEFEV 299  
QY 40 -----N-----GAL--V--P--P--AAL-----I--S----- 51  
Db 300 RPSWPLYLSANVGIOVLLFCTGAIRNEVTVDPVLVLPDVAAYCPAVVVKSVPGCEGVAL 359  
QY 52 II-Q-K-----G--LOY--V--E--AEV--S--I-----N----- 67  
Db 360 VQAQTKAISGLTIDGHAVGAKLTWEAVPGSEFSYAEVELGTADMIHTAETTLGLLTG 419  
QY 68 --E-----D-G-T-L----- 72  
Db 420 LAKAIGYATAADCGRTVLSPEPSCGMCQAACQRCQVVGKAGCAEASTAVCAQGDPH 479  
QY 73 -----FDGR-----P--IE----- 79  
Db 480 YTTFDGRRYDMGTCSYTMVELCSEDDTLPAFSVEAKNEHRRSVSYVGLVTVRAYSHS 539  
QY 80 -SLS-----LID-----AV-----M 88  
Db 540 VSLTRGEVGFVLVDNQRSLPVSLSERLRYQSGPRAVVVELVFLVTVTYDWDQALSL 599  
QY 89 P-----D-V-----V-----Q----- 93  
Db 600 PARFQOVCGLCGNYNGDPADDELTDPGALAPDAVEFASWKLDDGDLCEDCQNCNCPA 659  
QY 94 -TR-Q-QAY---R-----D-KLA-----Q-----Q----- 106  
Db 660 CTFGQAQHYEGDRLCGMLTKLDGPFPAVCHDTLDPFPFLEQCVYDLCVVGGERLSLGRGLS 719  
QY 107 --QA-----A-----A-----AAA----- 114  
Db 720 AVAAQACLELGISVGDWRSPANCPLSCPANSRYELCGPACPTSCNGAAAFSCNCRPCEG 779  
QY 115 -----A---A---AAS---Q-Q-G-S-A-K-----NGSN---T----- 130  
Db 780 CVCCLPGFVAGGACVPASSCGCTFGQLQAPGQEWADDELQORRCTCNGATHQVTCEDKQ 839  
QY 131 -----ANGE---EN-----GA--H-----T---I---AN- 143  
Db 840 SCFA-GERCSVQGLLGVCYPRFGTCQSGSDPHVSGFGRFRFDMGTCTVLLVSGCQNA 898  
QY 144 -----N-H-----T-----DMME----- 150  
Db 899 ALPAFRLVENEHRGSGTQSVSYTRAVREARGVKAVRREYPGQVLDVLLQYLPFOAAG 958  
QY 151 -----V-D-G-----D-----V--E----- 156  
Db 959 QVQVFRQGRDANVVRTDFGLTVTYDNNARVTAKVPSSYAEALCGLCGNFNGDPADDLALRG 1018  
QY 157 -----I-----P--P-----N----- 160  
Db 1019 GGOAANALAFGNSWQBETRPCCGATEPCPKLDSLVAQOLQSKNECIGILADPKGPPREC 1078  
QY 161 --K-----AV---VL-R----- 166  
Db 1079 HSKLDPQGAVRDCVYDRCLLPQSGPLCDALATYAAACQAAGATVHWRSEELCLPLSCPP 1138  
QY 167 -----C-----HE----- 169  
Db 1139 HSHYACSVGCPLSCGDLFPVPGCGSECHGCVCDGEPALSGESCLPLASCGCVHOGTYH 1198  
QY 170 -----S-----E-----V----- 172  
Db 1199 PGQTFYPGPCDLSLCHCEGGLVSCSSCGPHEACQPSGSGSLGCVAVGSSSTCOASGDP 1258  
QY 173 -----F-----I---C-----AW-----N- 178  
Db 1259 HYTTFGRRFDMGTCTVYVLAQTCTGTRPGLHRAVLEQNVAMNGRVSVTRVITVQVANF 1318  
QY 179 -----PVS-----D--L---LASG---S--G-D---ST-----A-----RI-- 195

Db	1319	TLRLRQWQKVTYVNGVDMKLPVVLANGQIRASQHGSDVVIETDFGLRVAYDLYVYRVTV	1378
Qy	196	-----P-----F-----H-----S-----	308
Db	1379	PGNYQQMGLCGNYNGDPKDDFKQKNGSQAGNANERFNSWEVVPDSPCLPPTPCPPGS	1438
Qy	197	-----P-----P-----ALD-----	313
Db	1439	EDCIPSHKPELEKKYQKEBEFCGLSSPTGLSSCHKLVDPQGPLKDCIFDLCLGGNL	1498
Qy	199	S-----H-----C-----I-----R-----	212
Db	1499	SILCSNIHAYVSAQAAGHVEPMRTETFCPMECPNPNHSHYELCADTCSLGCALSAPPQC	1558
Qy	208	-----Q-----L-----L-----R-----	212
Db	1559	QDGAECQCDGFLYNGQACVPIQQCGYHYVYEPETVLIIDNCRQOCTCHAGKGMV	1618
Qy	213	-----H-----C-----I-----R-----E-GGQ-----	220
Db	1619	COEHSCKPGQVQPSGGILSCVTKDPCHGVTCTPQETCKEQGGVCLPNEYATCWLWGD	1678
Qy	221	-----D-----D-----V-----P-----S-----	224
Db	1679	PHYHSPDRKDFQGTCTNYVLTATGCPGVSTQGLTPFTVTTKNQNRGNPAVSVRVTVTA	1738
Qy	225	-----N-----KD-----VT-----SLOWN-----	234
Db	1739	ALGTNISIHKDEIGKVRVNGVLTALPVSVDAGRISVTQGASKALLVADFLQVSYDWNMR	1798
Qy	235	-----S-----S-----E-----	236
Db	1799	VDVTLPSHYHAGVCLGNGMDRNPNDQVFPNGTLAPSIPIWGGSWRAPGWDPCLWDECR	1858
Qy	237	G-----T-----L-----LA-TG-----	243
Db	1859	GSCPTCPEDRLQYEGVGFGLAPGTGGPFTTCHAHVPPSPFKGVCGLDVMCGGDRDI	1918
Qy	244	-----SY-----	245
Db	1919	LKALASYAACQAAGVIEDWRAQVCEITCPENSHYEVGPPCPASPAPLTPPAV	1978
Qy	246	-----D-GF-A-R-I-----W-T-----K-DGN-----	257
Db	1979	CEGPCVGCQDAGFVLSADRCVPLNNGCGCWANGTYHEAGSEFWADGTCSQWCRCPGG	2038
Qy	258	L-----AS-----TL-----GOHKG-----PIF-ALK-----W-----N-----K-----	276
Db	2039	GSVLCTPASGLGEVCLLPSGGH-GQPVSTA-ECQAWGDHPHYVTLGDHGRFNGTCEY	2096
Qy	277	K-G-----NF-----I-----LSA-----	286
Db	2097	LLSAPCHGPPLGAENFTVTVANHRGSQAVSYTRSVTLQIYNHSLTLRARWPRKLOVDGV	2156
Qy	287	-----DK-----TT-----	290
Db	2157	FVTLPPFOLDLSLHAHLGADVVVTTTSGLSLAFDGDSEFVRLRVPAAYAGSLGCLGNYNQ	2216
Qy	291	-----H-----N-----KE-----I-YT-----I-----KW-----	399
Db	2217	DPADDLKAVGKPGAWQVGAQCGECVSKPCPCTPEQOESFGGPDAGVISATDGPL	2276
Qy	293	-----S-P-----T-----G-----P-----GT-----N-N-----	408
Db	2277	APCHGLVPPAQYFQGLLDACQVQGHGGCLCPAVATYVAAQAAGALREWRPDPFCPFQ	2336
Qy	295	-----AHT-----G-E-A-K-----Q-----	302
Db	2337	CPAHSHYELCGDSCPGSCPSSLAPEGESACRGCVCDAGFVLSGDTCPVPGQCGCLHDD	2396
Qy	303	-----Q-F-----	304
Db	2397	RYVPLGQTFYPPGCGDSLCRCREGGEVCEPSSSCPHETCTRPSGSLGCVAVGSTTCCQAS	2458
Qy	305	-----P-----F-----H-----S-----	308
Db	2457	GDPHYTTDFRRRFDPMGTCVYVLAQTCGTRPGLHRPAVLQENVANGNRVSVTRVITVQV	2516
Qy	309	A-----P-----P-----ALD-----	313
Db	2517	ANFTLRLEORQWKVTYVNGVDMKLPVVLANGQIRASQHGSDVVIETDFGLRVAYDLYVYRV	2576
Qy	314	V-----V-----D-----W-----	316
Db	2577	VTVPGNYQLMCLCGNYNGDPKDDFKQKNGSQAGNANERFNSWEVVPDSPCLPPTPCP	2636
Qy	317	-----Q-----Q-----	317
Db	2637	PGSEGCIPSEECPELEKKYQKEBEFCGLSSPTGLSSCHKLVDPQGPLKDCIFDLCLGG	2696
Qy	318	-----SN-----N-TF-----A-S-CS-T-----	327
Db	2697	GNLSILCSNIHAYVSAQAAGHVEPMRTETFCPMECPNPNHSHYELCADTCSLGCALSAP	2756
Qy	328	-----D-----MC-I-----H-----V-----C-K-----	335
Db	2757	LQCPDGAECQCDGFLYNGQACVPIQQCGYHYGAYYEPETVLIIDNCRQOCTCHAGK	2816
Qy	336	L-----Q-----D-----RP-----I-----	342
Db	2817	VVVCOEHSCKPGQVQPSGGILSCVTKDPCHGVTCTPQETCKEQGGVCLPNEYATCWL	2876
Qy	343	-----K-----T-F-----Q-----	346
Db	2877	WGDPHYHSPDRKDFQGTCTNYVLTATGCPGVSTQGLTPFTVTTKNQNRGNPAVSVRVV	2936
Qy	347	-----G-----HTNE-----VN-----A-I-----K-----	356
Db	2937	TVAALGTNISIHKDEIGKVRVNGVLTALPVSVDAGRISVAQASKALLVADFLQVSYDM	2996
Qy	357	W-D-----PT-----GNI-----LA-----S-----	366
Db	2997	NWRVDVTLPSSHYHAGVCLGNGMDRNPNDQVFPNGTLAPSIPIWGGSWRAPGWDPCLWD	3056
Qy	367	C-S-----D-----D-----D-M-----	371
Db	3057	ECRGCPTCPEDRLQYEGVGFGLAPGTGGPFTTCHAHVPPSPFKGVCGLDVMCGGD	3116
Qy	372	TL-K-----I-W-----SM-----	378
Db	3117	HDILKALASYAACQAAGVIEDWRAQVCEITCPENSHYEVGPPCPASPAPLPT	3176
Qy	379	K-Q-----D-----N-C-----VH-----D-----	386
Db	3177	PAVCEGPCVGCQDAGFVLSADRCVPLNNGCGCWANGTYHEAGSEFWADGTCSQWCRG	3236
Qy	387	-----L-----Q-----Q-----	389
Db	3237	PGGSLVCTPASGLGEVCLLPSQHGQCPVSTAECAQWGDHPHYVTLGDHGRFNGTCE	3296
Qy	390	-----H-----N-----KE-----I-YT-----I-----KW-----	399
Db	3297	YLLSAPCHGPPLGAENFTVTVANHRGSQAVSYTRSVTLQIYNHSLTLRARWPRKLOVDG	3356
Qy	400	-----S-P-----T-----G-----P-----GT-----N-N-----	408
Db	3357	FVTLPPFOLDLSLHAHLGADVVVTTTSGLSLAFDGDSEFVRLRVPAAYAGSLGCLGNYN	3416
Qy	409	-----PNA-NLM-----L-----A-----SAS-----	419
Db	3417	QDP-ADDLKAVGKPGAWQVGAQCGECVSKPCPCTPEQOESFGGPDAGVISATDG	3475
Qy	420	-----F-----D-----ST-V-----RLW-----	427
Db	3476	PLAPCHGLVPPAQYFQGLLDACQVQGHGGCLCPAVATYVAAQAAGALREWRPDPFCP	3535

428 QY -----D-----V----- 429  
3536 Db FQCPAHSHYELCGDSCPGSCPSLSAPGCSACREGVCVCDAGFVLSGDTCPVPGQCCLH 3595  
430 QY -----G-I-----CI----- 435  
3596 Db DDYYPPLGQTFYPPGPGCDSLCRCKREGGEVCEPSSCGPHETCRPSGSLGCVAVGSTCQ 3655  
436 QY -----H-----T-----L-----TK-----H-----OE----- 443  
3656 Db ASGDPHYTFDGRREFMGTCVYVLAQTCGTRPGLHRFAVLQENVAWNGRVSVTRVITV 3715  
444 QY -----PV----- 446  
3716 Db QVANFTLRLEQRQKVTNGVDMKLPVVLANGIRASQHSDDVVIEDFGLRVAYDLVY 3775  
447 QY -----SV-----G-I-----A-F-S-----PD----- 453  
3776 Db VRVTPGNYQLMCGLCGNYGDPKDDFKPNGSQAGNANEFNGSMEEVVPDSCPPLPPT 3835  
454 QY -----G-----RY-----LAS-G-----S-----PD----- 463  
3836 Db CPPSEGCIPSECPPELEKYYKEEFCGLLSPTGFLSSCHKLVDPQGLKDCIFDLCL 3895  
464 QY -----K-C-----V-----HI-----W-----N-----T----- 471  
3896 Db GGNLSILCSNIHAYVSACQAAGHVEPWRNETFCPMECPQNSHYELCACTCSLGSALS 3955  
472 QY -----Q-V-----CLH-----Y-----L-----N----- 479  
3956 Db APLQCPDGAEGCQDGSGLYNGACVPICQCGYHNGVYEPETVLIDNCRQOCTCHV 4015  
480 QY -----GOV-----LL-----N-----LG-----R-----S-----ICL-Y-T- 494  
4016 Db GKVVVCOEHSCKPQGVQCPGSGILSCVNDKPCPHGVTCRPOETCKEQGGQGVCLPNYEATC 4075  
495 QY -----PH-H-----LV-----V-----I-----P-----LV- 505  
4076 Db WLWGDPHYHSDGKPKDFQTCNVLATTCGPGVSTOGLTFPTVTKNQNRGNPAVSIVR 4135  
506 QY -----AL-----I-----E-----L-----L-----V-----L-----K 514  
4136 Db VVTVAALGTNISIHKBIGKVRVNGVLTALPVSADGRISVAQGASK 4182

RESULT 11  
US-08-471-119A-2  
; Sequence 2, Application US/08471119A  
; Patent No. 5827706  
; GENERAL INFORMATION:  
; APPLICANT: Leitner, Ernst  
; APPLICANT: Schneider, Elisabeth  
; APPLICANT: Schoergendorfer, Kurt  
; APPLICANT: Weber, Gerhard  
; TITLE OF INVENTION: Cyclosporin Synthetase  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5827706artis Corporation  
; STREET: 59 Route 10  
; CITY: East Hanover  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07936  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: /US/08/471,119A  
; APPLICATION NUMBER: US/08/471,119A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
; NAME: Kassenoff, Melvyn  
; REGISTRATION NUMBER: 26,389  
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 201 503 8474  
; TELEFAX: 201 503 8807  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15281 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Tolypocladium niveum  
; STRAIN: ATCC 34921  
; US-08-471-119A-2

Query Match 59.4%; Score 2533.4; DB 2; Length 15281;  
Best Local Similarity 7.7%; Pred. No. 1e-28;  
Matches 461; Conservative 46; Mismatches 4; Indels 5466; Gaps 373;  
QY 1 M-----S-----ISSDE-----VNF-----L-----V 12  
Db 5261 MKATNELSYRYAAVLHI-SDEPLPIYKIDPEAWINFEGSRLTREALAQVLKENENAESV 5319  
QY 13 -----Y-----R-----YL-----Q----- 17  
Db 5320 AISNIPYSKTVVERHIVRSIQEDANAPESMDGSDWISAVRTAQOCHTSLASDLFDIA 5379  
QY 18 E-SGP-----S-----HS-----A-F-----F----- 28  
Db 5380 EDAGFRVSVWARQHSQHSGALDAVPHHLKPATEDSRVLKFPDTHOGRPLKSLTNQPLP 5439  
QY 29 -----G-----I-----N-----GA----- 42  
Db 5440 AQSRRAELLIREGLQTLPPYMIPTSLIDRPLNANGKVDREARRAKITQSKPVE 5499  
QY 33 -----H----- 35  
Db 5500 DIVPPRNSVEATVCKGFTDVLGVVEGITDNFFNLGHSMLATKLAARGLQNLTRISVRD 5559  
QY 36 -----OSN-----I-----N-----GA----- 42  
Db 5560 VFDQPVVADLAAVIQRSAPHEPIKPADYTGVPVQPSFAQGRMLFQDLNNGVATWYLMPLG 5619  
QY 43 -----L-----V-----P-----P----- 46  
Db 5620 IRLHSLRVDALATAISALEQRHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKD 5679  
QY 47 -----A-----ALI-----S-----II----- 53  
Db 5680 AYLAVLKHEQTLFDLATEPGMRVALIRLGEERHILSIWMHIIISDGSVEVLFDENHRF 5739  
QY 54 -----Q-----K-----CL-----QY-----V-----E-----AEV----- 64  
Db 5740 YSSALRQDPMEOILPFIQYRDFAAWQKTEQVASHQRQLDYVTEHLADSTPAELLTDL 5799  
QY 65 -----SI-----N-----ED-----G 70  
Db 5800 PRESIISGRANELPTIEGLHDKLRAFCRVHQATPFVILLAAALRAAHVELTCAEDATLG 5859  
QY 71 T-----L-----F-----D----- 74  
Db 5860 TPIANRNPENLENNIGFFVNTQCMRIAIBENDONFESLVRVRSTATSAFANQVVPFESIV 5919  
QY 75 -----G-----R-----P-----I-----E-----S----- 80  
Db 5920 SSLLPGSRDASRNPLVQVILAVHVSQODLGLKLTLEGLRDEAVDSAISTRFDFVEHFLFEHAD 5979

QY 81 -LS-----L-I-----D-----ANMP-----DV 91  
DB 5980 RLGSVLYAKELPKLRTISVSVFLETLRRALDQPLTPLAVLPLTDGVEIASKGLD 6039  
QY 92 -----V-Q-TR-----QO----- 97  
DB 6040 PRTDYPRDANIVEFOOHRVATPDATAVKDATSILTYAQDQOSDLATLWLSRRHMPET 6099  
QY 98 -----AY-----R-----D-----KL-----A- 104  
DB 6100 LVGLAPRSCETIIFGIMKANLAVLPDINSAPARLSILSAVDGNKLVLGSGVTAP 6159  
QY 105 -Q-----Q-----Q-A-A-A-----A----- 110  
DB 6160 EQENPEVEAVGIEILLAGTGLDKTQSGNARPSATSLAYIFTSGTGKPKGVMEHRSVT 6219  
QY 111 --A-----A-A-A-A-A-A-----A----- 118  
DB 6220 RLAKPSNVISKLPOGARVAHLANAFDASIWEIATTLNGATLVCLDYHTVLD CRTKEV 6279  
QY 119 -S-----Q-----Q-----Q----- 121  
DB 6280 FERESITVTLMPALLKQCVAEIPETLAHLDLLYTGDRVGGHDAMRARSVLKIGMFSGY 6339  
QY 122 G-----S-----A-----K-----N-----GE----- 128  
DB 6340 GPTENTVISTIEVDADMPVNGPIGKTVNSGAYVMDRNOQLVPSGVVGLVVTGDGL 6399  
QY 129 -----N-----TANGE----- 134  
DB 6400 ARGYDPSLNKNRFIYITVNGESIRAYRTGDRVRYRPHDLQIEFFGRMDQVKIRHRIE 6459  
QY 135 -E-----N-----CA----- 138  
DB 6460 PGEVESALLSHNVQDAAVVICAPADQDSGAEMVAFVAARNTDEDEDTQEEBAVDQVGWE 6519  
QY 139 -H-T-----I-----ANN-----H-----T-----D----- 147  
DB 6520 THEPETAAYSEVKDIRQSEVNGDFMGWTSMDYDSEIDKTDHWEHLNDTMRMILDAREPGHV 6579  
QY 148 -----M-M-----E-----V-----DG-----D 154  
DB 6580 LEITGTGMVFNLAKECPGLQGYVGPEPSKAAQFVNDAAQSPALKDGRSIVHVGTATD 6639  
QY 155 -----VE-----IP----- 158  
DB 6640 INKAGPIQPLRVINSVAQYFPPEYLFVRVVEALVQIPSVRIVFGDMRTNAINRDFVAS 6699  
QY 159 -----P----- 159  
DB 6700 RALHTLGEKANKELVRQMIYELEANEELLTDPAFTSLRTRLGELKIKHVEIILPKTMKAT 6759  
QY 160 N---K---AVL---RG---H-----E----- 169  
DB 6760 NELSKYRYAAVLHVRGSRQESTHQVSPNAWIDFAADGLDRQTLINLLKHKDAGTVAIG 6819  
QY 170 -----SE-V-FI-----C----- 175  
DB 6820 NIPYSKTIVERFVYNKSLSEDDMEEGONSLDGSAAVAAVMAAQCPSLDMVDKEIAQEA 6879  
QY 176 -----A-W-N-----P-----V-----SD----- 182  
DB 6880 GYQVEVSWARQWONGALDAIFHHFFPKEGARTLIEFTDYEGRVNVTILNRPNSIQS 6939  
QY 183 -----LL-----A-----S----- 186  
DB 6940 RRLGTQIREKQLTLLPPYIMPSRIMVMDQPVNNGKIDRKELVRAIVAPKPSRAATRV 6999  
QY 187 -----G-S-----G----- 189  
DB 7000 APRNEIETAILRDEFDVLGTEVSLVONFFDLGGLSLMATKLAARVARRLDHAHISIKDVF 7059  
QY 190 -----D-----S-----T-----A-RIW-----NL----- 198

DB 7060 QPVLADLAASIQRESAPHEPIQRPYTPGAEQSPAQRLWFLDQLNLGATWYMLPIAIRI 7119  
QY 199 ----- 198  
DB 7120 RQQLRVAALSAAALPALERRHETLRTTFEESDGVGVQIVGEARNSDLRVHDVSTGDDGEYL 7179  
QY 199 -----SE----- 200  
DB 7180 EVLRREQTVPFDLSSEPGWRVCLVKTGEEDHVLISVMHHIYDGVSWVDILRGELGFYSA 7239  
QY 201 -----N-----ST----- 203  
DB 7240 ALRGODPLLHANPLFIQYRDFAAWQREAKQVEHQRLGYWSKQLVDSTPAELLTDLPRP 7299  
QY 204 --S--GS--T-----Q-----L-VLR--H----- 213  
DB 7300 SILSRAGSVDTIIGSVVYGALQSCFTRSVTTFVVLVTVFRIAHFRLTAVDDATIGTPI 7359  
QY 214 -----C-I-----R-----BG----- 218  
DB 7360 ANRNRPELTGVGCFVNTQCMRISIADDDNFEGLVQVRNVATAAYANQDVPPERIVSAL 7419  
QY 219 --G-----Q-----D-----D-----VP----- 223  
DB 7420 VFGSRNTRNPLVQLMFAVQSVEDYDQVRLEGLSVMMFGEASTRFDMFHLVPGDQKLT 7479  
QY 224 -----S-----N-----K-----D-----V-----TS 230  
DB 7480 GSVLYSSDLFEOGTQNFVDIFQECULRSVLDQPLPISVLPFSNAISNLESLLLEMPST 7539  
QY 231 -----GS-----LD-----W----- 233  
DB 7540 DYPRDRTVVDLFREQAAICPDSIAVKDSSQLTYAQLDQSDRVAWLHERHMPAESLVG 7599  
QY 234 -----N-----S-EG----- 237  
DB 7600 VLSPRSCETIIAYFGIMKANLAYPLDVYAPDARLAAILDTVEGERLLLLGAGVPOPGIQ 7659  
QY 238 --T-----L-----LA-----TG----- 243  
DB 7660 IPRLSYATAEALSHATTVDTVTSIPOPSATSLAYVIFTSGTGKPKGVMEHHRGIVRLVR 7719  
QY 244 -----S-----Y----- 245  
DB 7720 DTNVNVFPGSALPVSHFESNLANDAAATWEIYTAVLNGTGVVICIDRTDMLDIAALNSTPR 7779  
QY 246 -----D----- 246  
DB 7780 KENVRAAFPTPAFLKQCLAETPELVANLEILHTAGDRLDPGDANLAGKTAAGGIFNVLGH 7839  
QY 247 -----G--F-----AR 250  
DB 7840 TENTAYSTFPYVVBGETFVNGVPGVGRGISNSHAYIIDRHQKLVPAVGMGELIITGQGVAR 7899  
QY 251 -----I-----W--T--K-----DG----- 256  
DB 7900 GYTDSALNKDRFVYIDINGKSTWSYTGDKARYRPDQGLEFFGRMDQMKIRGVRIEPC 7959  
QY 257 ----- 256  
DB 7960 EVELTLDDHKSVLAAATVVVRPPNGDPEMIAFITIDAEDDVQTHKAIYKHLQILPAYMI 8019  
QY 257 -----N-----LA-----ST----- 261  
DB 8020 PSHLVLDQMPVTDNGKVDKDLALRAQTVQKRSTAAARVPRDEVEAVLCEEYSNLEEV 8079  
QY 262 -----LG-----Q-----HK----- 266  
DB 8080 EVGITDGFDLGHSLLATKLAARLSQNLTRVSVKDVDFDQPLADLADIIRGSHRHP 8139  
QY 267 -----GPI-----FA-----LKW-----N-----KKG-----N----- 279

Db 8140 IPATPYTGPVEQSPAQGL-WFLEQLNGASWYLMPEAIRMRGPLOTQKALAVNALVHR 8198  
Qy 280 -----F-----I-LS-AGVD-----K--TT----- 290  
Db 8199 HEALRTTTFEDHDGVGVQVIOFKSSQDLRIIDLSA-VDDTAYLAALKREQTTAFDLTSEP 8257  
Qy 291 -----II---W----- 293  
Db 8258 GWRVSLRLGDDDDYILSVMHHIISDGTVDVLRQELGQFYSAAIRGQEPQLSAKSLPIQ 8317  
Qy 294 --D----- 294  
Db 8318 YRDPFVWQEQENQIEQAQKLYWSQQLADSTPCPEFLTDLPRPSILSGEADAVPMWIDGT 8377  
Qy 295 -----AH--TG-----E----- 299  
Db 8378 VYQLLTDRCRTHQVTSFVLLAAPTAAHRTAHYRLTGTLDATVGTPIANRRNPELEGLIGFFVN 8437  
Qy 300 -----K--QO-----F--PF----- 306  
Db 8438 TQCRMAISETETESVQVRLTTTEAFANQDVPEQIVSTLLPGSRDTSRNPVQVMF 8497  
Qy 307 -----H-----S----- 308  
Db 8498 ALQSQDGLGRIQLEGMTDEALETPLSTRLDLEVHLFOEVGKLSGSLIYSTDLPEVETIRG 8557  
Qy 309 -----A-PAL-----D--V-- 314  
Db 8558 IVDVLEILRGLQPKORLMAMPITDITKLDRDQGLLTAKPAYPRESSVIDLFRQOVA 8617  
Qy 315 --D--W-----OSN-----N--T--FA--SC----- 325  
Db 8618 AAPDAIAVWSSSTLTADLDGQSNKLAHLWCORNMAPETLVAVFAPRSLTIVAFGLVL 8677  
Qy 326 ----- 325  
Db 8678 KANLAYLPDVNAPARIEALSAVPGHKLVLVQAHPGLTMADELTQIDEALASS 8737  
Qy 326 -----ST-----D-----M-- 329  
Db 8738 SGDEQIHASGPTATSLAYVMFTSGTGKPKGVMIHRSIIRLVKNSDVVATLPTPVMA 8797  
Qy 330 -----C-----I-----HV---C 334  
Db 8798 NVSNLAFDISVQEIYTALLNGTTLVCLDYTLTLDKILYNVVFVEAQVNAAMFTPVLLKQC 8857  
Qy 335 -----KL-----G--QD----- 339  
Db 8858 LGNMPAIIISRLSVLFNVGDRDLDAHDAVAASGLIQDAYVYNAVYGTENGQMSTMYKVDVNEP 8917  
Qy 340 -----R-----P----- 341  
Db 8918 FVNGVPIGRSITNSGAYMVGNOQLVSPGVMEIVVTGDLGARGYDTSALDEDRFVHVTI 8977  
Qy 342 -----IK--T-----P--O-----GH----- 348  
Db 8978 DGEENIKAYRTGDRVRYRPKDFEIEFFGRMDQVKIRGRIEPAEVEHALLGDLVHDA 9037  
Qy 349 -----TN-----E-V--N-----A-IR----- 356  
Db 9038 VVLRKANQEPMEAFITTSQDEETIEQESNKQVQGWGEHFDVSRVADIKDLDTSTFGHD 9097  
Qy 357 -----W-----D--P-----TG--NL----- 363  
Db 9098 FLGWTSMYGDVDPVNEKWEKDETTASLLDNRPFGHILEBIGAGTGMILSNLKGVDGLQK 9157  
Qy 364 -----LA----- 365  
Db 9158 YVGLDPAPSAAI FVNEAVKSLPSLAGKARVLVGTALDIGSLDKNEQPELVINSVAQYF 9217  
Qy 366 --SC-----S-----D----- 369  
Db 9218 PTSEYLLIKVVKVVEPSVKRVFFGDIRSQALNRDFLAARVRLGDNASKQIREKIAE 9277

Qy 370 -----D-----M--T-----L-----K-- 374  
Db 9278 LEESEBELVDPAFFVLSRSQLPNIKHVEVLPKMKATNELSSYRYAAVLHISHNEEQ 9337  
Qy 375 -----I--W-----S-----M-----K-----QD-- 381  
Db 9338 LIQIDPTAWDPFAATQKDSQGLRNLLQOGRDDVMIAVGNIPYSKTI VERHINSLDQDH 9397  
Qy 382 ----- 381  
Db 9398 VNSLDGTSWISDARSAAICTSPDAPALTQAKEGFRVELSWARQSONGALDAVFHRL 9457  
Qy 382 ---NC---VH----- 385  
Db 9458 ATDANCERSRVLVHFPTDHQRLRTLNRPLQARSRRIESQVFEALOTALPAYMIPSR 9517  
Qy 386 ----- 385  
Db 9518 IIVLPOMPTNANGKVRKQLARRAQVAVKRAVSARVAPRNDTEIIVLCHEYADILGTEVG 9577  
Qy 386 -----DL----- 387  
Db 9578 ITDNFDMGSHLMATKLAARLSRDLTRVTKEVPDKPLADLAASIEQGSTPHLPAS 9637  
Qy 388 ----- 387  
Db 9638 SVYSGPVEQSYAQRLWFLDQFNLNATWYHMSLMLRLLGFLNMDALDVALRLEQRHETL 9697  
Qy 388 ---Q---Q--H-----N--KE----- 393  
Db 9698 RTTFAQKDIGQVQVHEAGMKRLKVLDSLDKKEKHEHMAVLENEQMPFTLASEPGMKHGL 9757  
Qy 394 -----I-----YT----- 397  
Db 9758 ARLGPTEYILSLVHHMFSDGWSVDILRQELGQFYSALRGRDPLSQVKPLPIQYRDFAA 9817  
Qy 398 --K-----W-----S-----P--T--GP----- 404  
Db 9818 WQKEAAQVAHERQALAYENQLADSTPGELLTDPRPQFLSGKAGVIVPTIEGVVEXKL 9877  
Qy 405 -----GT--N--N--P-----N----- 410  
Db 9878 KPSKEROVTLPFVLLTAPRATHRPLTGAEDATIGTPIANRNPPELHIIIGFFVNTQCMRL 9937  
Qy 411 -----A--N-----LM----- 414  
Db 9938 LLDTGSTFESLVQHVRSVATDAYSNODIPFERIVSALLPGSRDASRPLIQLMFALHSQP 9997  
Qy 415 -----L--A-----S--A-----S--F-- 420  
Db 9998 DLGNITLEGLEHERLPTSVATFRDMEFHLQBPNKLSGILFADELFPQETINSVVTVFQ 10057  
Qy 421 -----D--ST-----V--R----- 425  
Db 10058 EILRGLDQPOVSIISMTPLTDGLIDLEKLGLEIESSNFRDYSVVDFVFRQVAAFNAP 10117  
Qy 426 -----L-----W----- 427  
Db 10118 AVVDSETSMYSYTLSDQKSEQIAAWLHAQGLRPESLTCVMAPRSFETIVSLFGILKAGAY 10177  
Qy 428 ---DV-----DR----- 431  
Db 10178 LPLDVNSPAARTQPIILSEVEGKRLVLLSGIDMPOSDRMDVETARIQDILTNKVERSOP 10237  
Qy 432 -----G--ICI-----H----- 436  
Db 10238 MSRPATSILAYVIFTSSTGRPKGVMI EHRNIIIRLVKQSNVTSQLPQDLRMAHISNAPD 10297  
Qy 432 -----G--ICI-----H----- 436  
Db 10298 ASIWEIFTAILNGGALICIDYFTLLDLSQALRTTTFEKARVNATLAFAPALLKECLNHPTLF 10357

QY 437 ----- 436  
Db 10358 EDLKVLYIGDRDLDAAKIQAIVKGTVMYVAYGPTENTVMSTIYRLTDGESVANGVPIG 10417  
QY 437 ----- 438  
Db 10418 NAVSSGAYIMQKRLVPPGVNGELVSGDGLARGYTNSLTNADRFVINDQKARAY 10477  
QY 439 ----- 444  
Db 10478 RTGDRTRYPKDGSIEFFGEMDQOVKIRGHRVPEAEVEQAMGLNKAHDAAVVQAVDQ 10537  
QY 445 ----- 447  
Db 10538 ETEMIGFVMSASDRFSEGBEETNQOEWDHPESTAYAGIEAIDQATLGRDFTSMY 10597  
QY 448 ----- 452  
Db 10598 NGNLIDKAEMEELDDTMQSLDKEDARPCAEIGTGTGMVLFNLPKNDGLESYVGIEPSR 10657  
QY 453 ----- 458  
Db 10658 SAALFVDKAAQDPGLQGTQIILVGAEDIKLVDFHPDVVVINSVAQYPPSRSYLVQIA 10717  
QY 459 S ----- 464  
Db 10718 SELIHMTSVKTIFFGDMRSWATNRDPLVSRALYTLGDKATKQIQROEVARLENEDELIV 10777  
QY 465 ----- 471  
Db 10778 DPAFTSLTSQWPKVHVILPKRMRTSNELSSRYAAVLHICRDGEGNRYGRVHSV 10837  
QY 469 ----- 471  
Db 10838 EENAWIDFASGMDRHALVQMLDERDRAKTVAGNIPHSNTINERHFTSLDTEGEGIAQ 10897  
QY 472 ----- 477  
Db 10898 DSLDGSAMQATKAMARCPCLSVTELVEIGQAAGFRVSVMARQSRQHGLDVVPHLE 10957  
QY 478 ----- 482  
Db 10958 DORVGRVLINFTDFERLPPSTGLTSRPLQRIQNRFRSQIRBQLQTLPLPYMPSRIVV 11017  
QY 483 L ----- 491  
Db 11018 LERMLNANSKVDKELAKARTLOTIKFSATRVAPRNDIEAVLDFEFOAVLGVTVGVMD 11077  
QY 492 ----- 497  
Db 11078 NPFELGHSILMATKLAARLSRLDTRVSVKDIQNPILQDLADVVQTSAPHEAIPSTPY 11137  
QY 498 ----- 501  
Db 11138 SGVPEQSFQRLWFLQNLNASHYHMLPLASRLGLRLEALQALATIEARHESLRTT 11197  
QY 502 ----- 514  
Db 11198 FERQDGVQVIVRAARNKQLRIIDVSGTEDIYLAALK 11234

## RESULT 12

US-09-919-497-70  
; Sequence 70, Application US/09919497  
; Patent No. 6773883  
; GENERAL INFORMATION:  
; APPLICANT: Mutter, George L.  
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER  
; FILE REFERENCE: B0801/7225  
; CURRENT APPLICATION NUMBER: US/09/919,497  
; CURRENT FILING DATE: 2001-07-31  
; PRIOR APPLICATION NUMBER: US 60/221,735  
; PRIOR FILING DATE: 2000-07-31

; NUMBER OF SEQ ID NOS: 100  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 70  
; LENGTH: 4861  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-919-497-70

Query Match 69.3%; Score 2527.7; DB 4; Length 4861;  
Best Local Similarity 9.1%; Pred. No. 2.8e-31;  
Matches 431; Conservative 60; Mismatches 19; Indels 4233; Gaps 365;

QY 1 M-S-I-----SS-D-----E-V-----N----- 9  
Db 1 MATMIPPVKLWLEHLNSSWITEDSESIATREGVAVLYSKLVSNKEVWPLPQOVLCCLKGP 60  
QY 10 -----F-----F-----LVY-R-Y-L----- 16  
Db 61 QLPDFERESLSDRDQDHYLDALLSSQLALAKMVCSDSPFAGALRKRLLLVLRVYFALS 120  
QY 17 -----Q--Es-----G-----F----- 21  
Db 121 YHDKGVKQQQHSPESSGSADVHVSERPRSTDALIEMGVTRGLSLLFALLRQSWMP 180  
QY 22 -----S-----H-----S-A-----F----- 26  
Db 181 VSGPGLSLCNDVIHTAIEVVSSLPPLSLANESKIPPMGLDCLSQVTTFLKGVTPNSGAD 240  
QY 27 TFG-----IK-----S-H----- 33  
Db 241 TLGRRLASELLGLAAQGRSLRYLLEWIEMLGASAVVHTMEKGKLLSQEGMISDFCFM 300  
QY 34 -I-----SQ-----SN-----I 39  
Db 301 TILMQRRSLGSSADRSQWRPTRTSDGLCSLYEAALCLFEEVCRMASDYSTRCASPD 360  
QY 40 -NG-A-----LV-----P-----P-----A-----L-----I-- 50  
Db 361 QTGDAPIVSETCEVYVWGSNSHQVTEGQEKILQPKLAPSFSDAQTIAGQYCTFVIST 420  
QY 51 -----SI-----I- 53  
Db 421 DGSVRACGKSGYGRGLGDSNNQSTLKLTFBPHRSIKKVSSSKSGSDGHTLAFTTEGEV 480  
QY 54 -----OK-----G-LQ-----Y-----V-----E-----AE--- 63  
Db 481 SWGDGDYKLGHSNSTQYPKLIQGLPQGVVVCVSAGYRHSAAVTEGELYTWGE 540  
QY 64 -----V-SI-N-E-----DG-TL--F--D-----G--- 75  
Db 541 GRLGHCDSNSRNIPTLVKDI SNVGEVSCGSSHTIALSKDGRVWVSFGGDNGLGHGDTN 600  
QY 76 R--P--IE-----SL-----S-----LI- 84  
Db 601 RVYKPKVIEALQGMFKVCAQSQSSALSTSTQVYVWGCACGLCGSSEATLRPKLIE 660  
QY 85 -----D--A-----V-----M-----P-----D-V-V 92  
Db 661 ELAATRIVDVIGDSHCLSHLDNHDNVEYAWGNNSMGCGGCGNSTGPTTKPKVSGLDGIAI 720  
QY 93 Q-----T--R-QQ--A--Y-----R--DK----- 102  
Db 721 QQISAGTSHSLAWTALPRDQVVAWHRPYCVDLBESTFSLHRSFLRYCDKINSEIPPLP 780  
QY 103 -----LA-----QQ-----QA 108  
Db 781 FPSREHSHFLKCLKLLSNHLALAGGVATSIILGRQAPLRLNLLFRLMDSTVPDEIOE 840  
QY 109 AA-----AAAAA-----AAS-----Q----- 121  
Db 841 VVIETLSVGATMLLPPLRERMELLHSLLPQGPDRWESLSKGORMQDILITSLQDTHVA 900  
QY 122 ---G---S-AK-----N-----G-----E-----N-----T--AN 132

Db 901 SLLGYSPSADADLSSVCTGVGNLSQDPYGTQSCHPDTHLAEILMKTLNLLNGFYTDQAF 960  
Qy 133 GE-B-N-----G-----AH-----T-----IAN---N-----144  
Db 961 GELEKNDKFLGTSSSENSOPAHHELLCSLQKQLLAFCHINNISENSSVALLKHHLQ 1020  
Qy 145 ---H-TD-----I-----MM-----149  
Db 1021 LLLPHATDIYSRSANLLKESPMNGSVGEKLRDVIYVSAAGSMCLQIVNSLLLPVSVARP 1080  
Qy 150 -----E-VD-G-----D 154  
Db 1081 LLSYLLDLLPLDCLNLLPAADLLEDEQLQWPLHGPELIDPAGLPQPAQSWWLVLD 1140  
Qy 155 VE-----I-----P-----159  
Db 1141 LERTIALIGRCGLGMLQSPVSPPEEQDTAYWMTPLFSDGVEMDTPLQDKMCSLLEVA 1200  
Qy 160 ---N-----K-----AV-----163  
Db 1201 LSGNEOKPFDYKLRPEIAVYVDLALGCSKEPARSLWISMODYAVSKDWDSATLSNESLL 1260  
Qy 164 ---VL-----R-G-HSEVP-----I-----C-----175  
Db 1261 DTVSRFVLAALLKHTNLLSQACGESRYQPKHLSSEVYCVKYSRLLACKNLELIQTRS 1320  
Qy 176 -A-W-----177  
Db 1321 SSRDRWISNQDSADVPQESHSTRTTIDEEAEEMEQAEARDREHGHPEDDEBERHEVM 1380  
Qy 178 -----N-----P-----VS-----D-----182  
Db 1381 TAGKIFOCFLSAREVARSRRDRMNSGAGGARADPPPOQOERRVSTDLPSGQVYTA 1440  
Qy 183 -----LL-----AS-----186  
Db 1441 ACNSVIHRCALLILGVSPVIDELQKREEQLOQPSTASGEGGLMTRSESLTAESRLVH 1500  
Qy 187 -----G-SG-----D-----S-T-A 193  
Db 1501 TSPNYRLIKRSBDSLQSPESDEGVALSGRONVDLDAASHRKRGPMHSLQESLSDSWA 1560  
Qy 194 RI-----W-N-----L-S-----EN-----STS-204  
Db 1561 RLKHSRDWLCNSSYFESDFLTYSKLGVHTLIENVVSFVSGDVGNAPGFKPEPESMSTSP 1620  
Qy 205 -----GS-----T-----Q 208  
Db 1621 QASTIAEQOQLRAELRLEALHQLVLLSGWEEKGSLAGSLSSGFSQSSTLLTSVRLQ 1680  
Qy 209 -----L-----V-----L 211  
Db 1681 FLAGCFGLTGVHTGAKGESRLHYODGIRAAKRNIQIBIQAVVHKIYQOLSATLERAL 1740  
Qy 212 -----R-----H-----C-----214  
Db 1741 QANKHHIEAQORLLLVTFVFSVHYQPDVVSLSAISTGLNLVLSQLCOTDTMLGQPLQLLP 1800  
Qy 215 -----I-R-----186  
Db 1801 KTGVSQSLTALKVASTRELLQILAITTGTYADKLSPKVQVQSLDLCLCSQLKNLLSQTGVLH 1860  
Qy 217 -----EG-----G-----Q-----D-----V-----222  
Db 1861 MASFGEGEQDBEEKKVDSSGETEKKOPRAALRKQHAELHGLDFLPLFRVSVSSKAI 1920  
Qy 223 -----P-----S-----N-----KD-----V 228  
Db 1921 QSKWAPKWTVEVLNLTASQKCSGILPLVGNRLTRLLALHVLBAVLFACEGSGVEDDQMAQI 1980  
Qy 229 T-----SL--D--W-----N-----S-----235

Db 1981 VERFLSLLSDCMWETPIAQAKHAIQIKEQEIKLQKQGELEEDENLPIQEVSPDEKA 2040  
Qy 236 ---E-G---T-----L-LA-TG-----243  
Db 2041 QCCLVENGOILTHSGSGKGYGLASTGVTGCVQWKFYIVKENRGNEGTCTGVSRWPVHDF 2100  
Qy 244 ---S--Y-----D-----G-----247  
Db 2101 NHRTTSDMMLYRAYSGNLYHNGEQTLTSSFTQGDFTICVLDMEARTISFGKNGBEPKLA 2160  
Qy 248 -----F-----A-----249  
Db 2161 FEDVDAEALYPCVMFYSSNPGEKVKICDMQMRGTPRDLPLGDPICSPVAALAEATIQLV 2220  
Qy 250 RI-----WT-----KD-G-----N-----257  
Db 2221 RILHRTWTYCIKKNOMERLHKIKICESGOKLKKRSVQSREBENREEMKEEKES 2280  
Qy 258 ---LA--S---TL-----G-----Q--H--K-266  
Db 2281 GKTRHGLADLSELQURTLCIEVWPVLAVIGVDAGLRVGGRCVHKQTRHATLLGVVKE 2340  
Qy 267 G-----P-FALKW-----N-----275  
Db 2341 GSTSAKVQWDEAEITISFPTF---WSPSDTPLYNLEPCPLPFDVARFRGLTASVLLDLT 2397  
Qy 276 ---KK-----G-----278  
Db 2398 YLTGVHEDMGKOSTRHEKHRESEBEGDVEQKPESESALDMRTGLTSDDKVKSQTSS 2457  
Qy 279 ---N---F-----I-----281  
Db 2458 KSENEIASFSLDPTLPSPVESQHOITEGKRKNHHSKNHDVAQSEIRAVQSLVLYLGAMK 2517  
Qy 282 -LSA-----284  
Db 2518 SILSALIGCSKYAELLILPKVLAENGHNSDCASSPVVHEDVEMRAALQFLMRHVMKRAVMR 2577  
Qy 285 ---G-----V--D--K-----288  
Db 2578 SPIKRALGLADLERAQAMITYKLWVHGLLEDQFGKIKOIBDQQAESDPAQQAQTPVTT 2637  
Qy 289 ---TT-----290  
Db 2638 PRASSTTSFMSSSLEDTTATTPTVDTETVPASESPCVMPPLSLLRQMFSSYPTTIVLPT 2697  
Qy 291 -----290  
Db 2698 RAQTPPISLPTSPSDEVGRQSLTSPDSQSARPANRTALSDPSSRLSTSPPPPAIAVPL 2757  
Qy 291 ---I-----I---W-----293  
Db 2758 LEMGFSRLQIAKAMEATGARGEADAQNITVLAMWMIHCPHEDEEPEQSGSTADSRPGA 2817  
Qy 294 ---DA-----HT-----GE---A--300  
Db 2818 VLGSGKSNDCPYLQSPGDIPSADAAEMEEGSESPDNLDHTENAASGSGPSARGSAVT 2877  
Qy 301 ---K-----QO-----303  
Db 2878 RRRKFDLAARTLLARAAGLYRSVQAHRNQRREGISLQDDPGALYDFNLDELEIDLDD 2937  
Qy 304 ---F-----P-----F-----306  
Db 2938 AMEAMFGDLTSDNLDILGMWIPVLDWPTWHVCESEDREVVVCELCECVSVFNQMKR 2997  
Qy 307 -H---SA-----P-----310  
Db 2998 NHPGCRSANROGRYNSGVYDGMFGEGCGSGNPYYLLCGTCKEYKAMTKSKTSSE 3057  
Qy 311 ---A--L---D-V--DW-----Q---317  
Db 3058 YKQAPDLIGKQDSVVEEDWMDLVDDEKLAGEBEFELLAGFLGLNDRRIIVPEPVQFPD 3117



QY 318 -----S-----N-----N----- 320  
Db 3118 SDPLGASVAMVATNSWEETLMQIGCHGSVEKSSGRITLGEQAALANPHDRVVALRV 3177  
QY 321 T-----FA-----SCS-----TD-----MCI----- 331  
Db 3178 TAAAOVLLARTVMWRALSLLSVSGSCSLAAGLESIGLTDIRTLVRLMCLAAAGLAGLST 3237  
QY 332 -----H-----V-CKL-----G-----O----- 338  
Db 3238 SP6MASTERSRGGHSHKANKPISC-LAYLSTAVGCLASNAFSAKLLVOLCTONLISAA 3296  
QY 339 -----D-R-P-I-K-T-F-Q----- 346  
Db 3297 TGVNLTVDSDIORKFLPSFLRGIAEBENKLVSPNFVWTOALVALLADKGAKLRPNYDKS 3356  
QY 347 -----G-----H-----T-----N----- 350  
Db 3357 EVEKGPLELANALAAACCLSSRLSSQHRQWAAQOLVRTLAHHRDQTTLOTILADMGDL 3416  
QY 351 -----E-----V-N-A-----IK-W----- 357  
Db 3417 RKCSFIKLEAHQNRVMTVCWCKKGLLATSGNDGTTIRVMNVTKQYSLOQTVCVFNLEGD 3476  
QY 358 -----DPT-----G-----N-----L-----LA----- 365  
Db 3477 ABESLSPSPSPSPVSWISGKYLALAKEMVINQVNGGKGLVDIQHWSALAWPEE 3536  
QY 366 -----S-C-S-D----- 369  
Db 3537 GPATANGSPPELLVGRMDGSLGLEVDVSTMRRELEHCYRKDVSVTCIAWFSEDRP 3596  
QY 370 -----D-----M-----T-----L-----K----- 374  
Db 3597 FAVGYFDGKLLGTBPEKGGIVLIDAHKDTLISMKWDPTGHILMTCAKEDSVKLGSI 3656  
QY 375 -----I-W-----S-M-----KQD----- 381  
Db 3657 SGCWCLLSHCHPSIVNGIANGTAWCLPCKGKQLLMATGOSGLVCVWRIPQDTTQTNVTS 3716  
QY 382 -----N-----CV-----H----- 385  
Db 3717 AEGWMDQESNCQGYRKSSGAKCYQLRGHITPVRTVAFSSDGLALVSGGLGLMNIWSL 3776  
QY 386 -D--IQ-----Q-----H-----N-KEIY-----TIKW-----S-- 400  
Db 3777 RDGVLQTVVIGSGAIQTTVWIPEVGVAAACNSRSDVNVNCTAEWAAAHVLTATCTAL 3836  
QY 401 -----P-T-G-----P-----GT-- 406  
Db 3837 KQGVGLNAPCMRAFLEPLPMLQEQVAYEPHVVGQDLVHSPYMQCLASLAVGLHL 3896  
QY 407 -----N-----P-----NAN-----L-----MLA-SA-----SF----- 420  
Db 3897 DQLCNPPVPPHQNCLPDPASNNPNEWALECFSTTIKAAEALITNGAQPFESFTVPDLE 3956  
QY 421 -----DSTVRL-----W-----D-V-----D-----R-----G-- 432  
Db 3957 VPPEDELVLMDNSKWMINGMDEQIMSWATSRPBDWHLGGKCVYLMWAGRHGQABAGRN 4016  
QY 433 -----IC-----I-----H----- 436  
Db 4017 VMVPAAPSPSQAQVICQNGCTFTVIQANGTVLACGEGSVGRIGQNSDDLHLVLTVISAL 4076  
QY 437 -----T-----LT-----K-H-----Q-EP----- 444  
Db 4077 QGFVTVLTVTSCSDGSHMALTSGEVFSWGDGDKLGHGNSDRQRPRQIEALOGEEV 4136  
QY 445 V-----Y-S--VA-----FS-----PD-GR----- 456  
Db 4137 VQMSCGFKHSVAVTSDGKLFTEGNGDYGRGLGNTSNKKLPERVTALEGYQIGQVACGLN 4196

QY 457 --LA-S--GS--F--D--K--C----- 465  
Db 4197 HTLAVSADGSMWAFGDGDKGLGNSTAKSSPOKIDVLGIGIKKACGTQFSVALTK 4256  
QY 466 ---V-----H-----I-----W- 469  
Db 4257 DGHVYTFQODRILGLPEGRARNHRPQIPVLAGVIEDVAVGAETLALASNDVYAMG 4316  
QY 470 -----N-----T-----QV-----C----- 474  
Db 4317 SNSEGOGLGHTNHVREPTLVTLTGLOKKNVROI SAGRCHSAANTAPPVPPRAGVSVPLQL 4376  
QY 475 -----Y-----LN----- 479  
Db 4377 GLPDTVPPQYALREVSIHTVRARLLYHFSDLMTYSWRLLNLSNNQNSTSHYNAGTW 4436  
QY 480 ---GQ-----VL-L-----NLG----- 487  
Db 4437 GIVQGLRPLAPRVYTLPMVRSIGKTMVOGKNYGPQITVKRISTRGRKCKPIFVOIARQ 4496  
QY 488 -----R-S-----I-C-----L----- 492  
Db 4497 VVKLNASDLRLPSRAWKVLVGEAGDAGVFDITTEMCOBLETGIVDLLIPSPNATAE 4556  
QY 493 --YT--L--P--H-----HL----- 499  
Db 4557 VGYNRDRFLFNPSACLDLHMQFKFLGILMVARTKPKDLHLAPLVKQCCVPLTLE 4616  
QY 500 ---V-----V-----IPL--V--A-----L-- 507  
Db 4617 DLEEDVLLVQTLNLSILHIEDSGITEESPHEMPLDSFVGQADGKRVPIIPGNSIPLT 4676  
QY 508 -----TE-----LL----- 511  
Db 4677 FNRKEYVERALEYRLHENDROVAAVREGMSHIVPVLSSLTAKOLEQMVCGMPSISVE 4736  
QY 512 VLK 514  
Db 4737 VLK 4739

RESULT 13  
US-08-460-751-2  
; Sequence 2, Application US/08460751  
; Patent No. 5891628  
; GENERAL INFORMATION:  
; APPLICANT: Reeders, Stephen  
; APPLICANT: Schneider, Michael  
; APPLICANT: Gluckmann, Sandra  
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY  
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Pennie & Edmonds  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,751  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/413,580  
; FILING DATE: 03-MAR-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Coruzzi, Laura A.

REGISTRATION NUMBER: 30,742  
REFERENCE/DOCKET NUMBER: 7638-005  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4303 amino acids  
TYPE: amino acid  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
US-08-460-751-2

Query Match 69.3%; Score 2527.4; DB 2; Length 4303;  
Best Local Similarity 10.0%; Pred. No. 1.5e-31;  
Matches 410; Conservative 82; Mismatches 14; Indels 3576; Gaps 355;

1 MSIS-----S-----D-----E-----EVN-----F-----10  
72 LDVSHNLLRALDVGCLLANLSALAEIDISNNKISTLEEGIFANLNLSEINLSGNPFECDC 131  
11 LVY-----R-----Y-----15  
132 GLAWLPQWAEQQVVRVPEAATCAGPGSLAGQPLLGIPLLDSCGGEYVACLPLDSSGT 191  
16 -----LO-----E-----15  
192 VAAVFSAAHEGLQPEACSAFCSTGQGLAALSEQGWCICGAAQPSASFACLSLCSGP 251  
21 -----F-----S-----H-----S-----AF-----T-----FG-----29  
252 PAPPAPTCRGFTLLQHVFPASPAGNLTGPHGPGSLAQALAFHIAAPLPVTDTRWDGDS 311  
30 --I-----KSH-----I-----S-----QSN-----38  
312 AEVDAAGPAASHRYVLPGRVHTAVLALGAGSALLGTDVQVEAAPAALELVCPSSVQSD 371  
39 -----I-----N-----G-----AL-----V-----P-----46  
372 SLDSIQNRGGSGLEAAYSIVALGEEPARAVHPLCPFSDTEIFFNGHCHYELVVEKAWLQ 431  
47 -----AA-----LIS-----I-----IO-----K-----G-----L-----57  
432 AQEQCQAWAGALAMVDSPAVQFLVSRVTRSLDVGWFGFTVQGVGVPAPQGEASLES 491  
58 -Q-----EDG-----YV-----E-----AE-----V-----S-----65  
492 CQNWLPGEHPATAEHCVRGLGPTGWCNTDLCSAHPSYVCELPQGGPVQDAENLLVGAPSG 551  
66 -----IN-----EDG-----TL-----F-----73  
552 DLQGLPLLAQDGLSAPHEFVEMVPPGLRLSREAFITTAETGTOELRRPQLRLQVYR 611  
74 -----D-----GR-----P-----I-----78  
612 LLSTAGTPENGSPESPDPNRITQLAPACMPGGRWCPCGANICLPDASCHPQACANGCTS 671  
79 -----E-----S-----LSLI-----DAV-----M-----P-----DVV-----92  
672 GPGGLGAPYALWREFLFSVPAGPPAQS-VTLHGQD-VLMLPGDLVGLQHDAGFGALLHC 729  
93 -----Q-----T-----R-----Q-----Q-----97  
730 SPAPGHGPRAPYLSANASSWLPHLPAQLEGTGCPACALRLLAQREQLTVLLGLRPNPG 789  
98 -----A-----Y-----RD-----KL-----A-----Q-----105  
790 LRLPGRVEVRAEYGVSRHNLSCSPDVSPVAGRLVIYPAPRDGRLYVPTNGSALVLQV 849  
106 -----QQ-----A-----A-----A-----A-----112  
850 DSGANATATARPWPGSLSARFENVCPALVATFVPACFWETNDTLFSVVALPMLSEGEHV 909

QY 113 -----A-----A-----A-----A-----A-----A-----118  
Db 910 DVVENSASRANLSLRTABEPCIGLRATPSPPEARVLQGVLYRSPVVEAGSDMVRWTI 969  
QY 119 -----S-----Q-----Q-----Q-----Q-----Q-----A 124  
Db 970 NDKQSLTFQNVFNVLYQSAAVFKLSLTASNHVSNVTVNVNVTVVERMNRQGLQVSTVPA 1029  
QY 125 -----KN-----GE-----N-----N-----TA-----131  
Db 1030 VLSPNATLALTAGVLVDSAVEAFLWTFDGEQALHQFPYPYNESFPVDPSPAQLVVEH 1089  
QY 132 N-----GE-----EN-----I-----ANNHT-----G-----A-----H 139  
Db 1090 NVHTYAAPGEYLLTVASNAPENLTQQVPVSVRASLPSVAVGVSDGLVAGRPVTFYPH 1149  
QY 140 -----T-----I-----ANNHT-----G-----A-----146  
Db 1150 PLPSPGGVLYTWDGDSPLVTQSQPAANHYSASRGTYHVRLEVNNTVSGAAAQADVRVF 1209  
QY 147 -----DM-----M-----E-----V-----D-----GD-----154  
Db 1210 EELRGLSVDMSLAVEQCAPVWVAATQTDGNTITWTFDMDGDTVLSGPEATVEHVYLAQN 1269  
QY 155 -----VE-----VE-----IP-----P-----NKA-----162  
Db 1270 CTVTVGAGSPAGHARSLLHVLVFLVLEVRPEAPACIPTQDARLTAYVTGNPAHYLFDWT 1329  
QY 163 -----V-----V-----L-----R-----G-----H-----ESEV-----172  
Db 1330 FGDSSNTTVRCGPTVTHNFTSGTFLALVLSRRVRAHYFTSICVEPEVGNVTLOPER 1389  
QY 173 -FI-----CAM-----NP-----VS-----D-----LL-----AS-----186  
Db 1390 QFVLGDEAWLVACAMPFPFYRYTWDFTGEAAPTREARPEVTFIYRDGSLVTVTASN 1449  
QY 187 -----GS-----GS-----GD-----190  
Db 1450 NISAANDSALVEQEPVLTSTIKVNGSLGLELOQPYLPSAVGRGRPASVYLDLGDGWL 1509  
QY 191 -----S-----TARI-----WN-----L-----SE-----N-----S 202  
Db 1510 GPETHAYNSTGDTFVRVAGMNEVSSEAWLNVTKRRVGRGLVWNASRTVPLNGSVSPS 1569  
QY 203 TS-----GS-----TOL-----VL-----RHC-----I-----RE-----G-----QD-----221  
Db 1570 TSLEAGSDVRYSNVLCDR-CTPIPGGPTISYTFRSVGTENIIIVTAENEVGSAQDSIFVYV 1628  
QY 222 -----V-----PSN-----KD-----V-----T-----SL-----231  
Db 1629 LQIEGLQVGGGRYPPTNHTVQLQAVVRDGTNVSVSWTAWDRDRGALLAGSGKFSLTVL 1688  
QY 232 -----D-----D-----W-----N-----234  
Db 1689 EAGTYHVQLRATNMLGSAWADCTMDPVEPVGMLWVAASPNPAAVNTSVTLASLAGSGV 1748  
QY 235 -----S-----EG-----T-----L-----LAT-----GS-----YD-----246  
Db 1749 VYTWLSLEGLSWETSPPFTTHSPPTPLGLHLVTMTAGNPLGSANATVEVDQVPVSGLSIR 1808  
QY 247 -----G-----F-----A-----R-----I-----W-----TK-----D-----G-----NL 258  
Db 1809 ASEPGGSFVAAGSVFPWQQLATGNTNVSWCNAVPGSSSKRGPHVTWVFPDAGTFSIRLN- 1867  
QY 259 AS-----T-----T-----L-----L-----Q-----H-----K-----G-----267  
Db 1868 ASNAVSWVSATYNTLTAEPIVGLVWASSKVVAQGLVHFQILLAAAGSAVTPELOVGGAN 1927  
QY 268 P-I-----F-----AL-----K-----W-----N-----KKG-----278  
Db 1928 PEVLPGPRFSGHPRVGRDHDVSVGRKNHVSWAQAVRIIVVLEAVSGLQVFNCCPEGIATG 1987

QY 279 --NF-----ILS-----AG----- 285  
Db 1988 TERNFTARVQGRVAVYAFSLQYQDGLVILSGRDVITYTVAAGLLEIQVRAFNALG 2047  
QY 286 --V-D-----K--T-----T-----II--WD----- 294  
Db 2048 SENRTLVLEVQDAVQVALQSGPCFTNRSQAFAATSPSPRRVAYHWFQDGSFGQDTDB 2107  
QY 295 --A-HT--G--E--A-----KO--Q-----FP----- 305  
Db 2108 PRAEHSYLRPGDYRVQNASNLVSFFVAQATVTVQVLACREPEVDVVLPLOVLMRRSQRN 2167  
QY 306 F--H-----SA-----P--AL--DV----- 314  
Db 2168 YLEAHVLDRCVITYQTEYRWEVYRTASQCRPGPARVALPGVDVSRPLVLPRLALPVGH 2227  
QY 315 -----D-----W-----QS--N-- 319  
Db 2228 YCFVVSFGDTPLTQSIQANVTVAPELVPPIIEGGSYRVMSTPDLVLDGSESDPNLE 2287  
QY 320 -----N-----TF----- 322  
Db 2288 DGDQPLSFHWACVASTQREAGCALNFGPRGSSVTIIPRERLAAGVEYTFSLTWKAGR 2347  
QY 323 --A-----SC----- 325  
Db 2348 KEEATQTVLIRSGRPIVLSLEVCVCKAQAVVEVSRSSVYVLEGRCLNCSGSKRGWAA 2407  
QY 326 --S--T--D-----M-----C--I----- 331  
Db 2408 RTFSNKTVLDETTTSTGSMRLVLRGVRDGEYTFTLVLRSGSEEGCASIRLSP 2467  
QY 332 -----H-----V-----C- 334  
Db 2468 NRPLGSCRLPFLGAVHALTTKVHPECTGWHDAEDAGAPLVYALLRLRCRQHCBEFCV 2527  
QY 335 -K-----LG----- 337  
Db 2528 YKGLSSYGAVLPPGPRHFEVGLVAVVQDLGAVALNRSIAITLPEPNGSATGLTW 2587  
QY 338 -----Q-D-----R-----P----- 341  
Db 2588 LHGLTASVLPGLLRQDPQHVIEYSALVTVLNEYERALDVAAEPKHQRAIRKNIT 2647  
QY 342 -----I-----K-T-----FO----- 346  
Db 2648 ETLVLSRVHTVDIIQIAAALAQCMGFSRELVCRCCLKQTLHLEAMMLILOAETTAGTV 2707  
QY 347 -----G--H-----T----- 349  
Db 2708 TPTAIGDSILNITGDLIHLASSDVRAPOPSSELGAESPSRWASQAYNLTSALMRLMRSR 2767  
QY 350 --NE----- 351  
Db 2768 VLNEEPLTAGEEIVAQGRSDPSRLCYGAPGPGCHFSIPEAFSGALANLSDVVQLIF 2827  
QY 352 -V-----N-----AI--K-----W----- 357  
Db 2828 LVDSNPPFGYISNYTVSTKVASMAFQTOGAQIPIERLASERAITVKVPNNSDWAARGH 2887  
QY 358 -----D--P--TG--N-----LLA----- 365  
Db 2888 RSSANSANSVVQPOASVGAVTLDSSNPAAGLHLQNLTYLLDGHVLSPEPEYLAVALH 2947  
QY 366 S-----CS-----D----- 370  
Db 2948 SEPRNEHNSARRIRPESLOQADHRPYTFTISPGSRDPAGSYHLNLSHFRWSALQVS 3007  
QY 371 --M-T-L-----KIW-S-----MK-----Q----- 380  
Db 3008 VGLYTSLCQYSEBDMVMWRTGELLPLEETSPROAVCLTRHLTAFGASLFPVPPSHRVFP 3067  
QY 381 ----D-N-----C--V-----H--DLQ-----QHN--K-EIYTIK 398

Db 3068 EPTADVNYITMLTCAVCLVTYVMAAILKLD-QLDASGRAIPFCQGRGRFYEL-VK 3125  
QY 399 --W-----S-PT-----G-----P--GT----- 406  
Db 3126 TWGRSGGTTAHVIMLYGVDSRSGHRHLDGRAPHRNLDIPRIATPHSLGSVWKIRVW 3185  
QY 407 -NN-----P-----N-----AN--L--M--L-AS-AS- 419  
Db 3186 HNKGLSPAFLQHVIVRDLQTAARSAPFLVNDWLSVETEANGGLVEKVEVLAASDAALLRF 3245  
QY 420 -----FD-----S--T-V-R-----L-----W-----D- 428  
Db 3246 RELLVAELQRFQKHIMLSIWDPRPSRFTRIQATCCVLLICLFLGANAVMYGAVGDS 3305  
QY 429 -----VD-----R-----G----- 432  
Db 3306 AVSTGHVSRSLSPSLVDTVAVGLVSSVVVYPVYLAIFLFRMSRSKVAGSPSPAGQOVL 3365  
QY 433 -I--C-----I-----H-----TLT----- 439  
Db 3366 DIDSCLDSSVLDSFLTTFGLHAEQAFVQMKSDLDLDDSKSLVWCWPSGEGTILSWPDLLS 3425  
QY 440 -----K----- 440  
Db 3426 DPSIVGSNLRQLARGAQHGLGPEEDGFSLASPYSPAKSFSASDEDLIOQVLAEGVSPA 3485  
QY 441 -----H-----Q-----E-- 443  
Db 3486 PTQDTHMETDLLSSISTPEKTETLALQRLGELGPPSPGLNWEOPQAAARLSRTGLVEGL 3545  
QY 444 -----P-----V--Y-----SVA----- 449  
Db 3546 RXELLPAWCASLAHGLSLLLVAVAVAVSGVGCASPPGVSVANWLLSSSASFLASFLGWEP 3605  
QY 450 -----FS-----PD----- 453  
Db 3606 LKVLLEALYFSLVAKRLHDEDDTLVESPAVTPVSARVPRVRPPHGFALFLAKEARKVK 3665  
QY 454 --G--R--Y-----LAS-GSPD-KC-----V--H-----I----- 468  
Db 3666 RUHGMRLSLLVMLFLVTLTASYG--DASCHGHAYRLOSAIKQELHSRAFLAITRSEEL 3723  
QY 469 --W-----N--T-----QV-----C-----L-- 475  
Db 3724 WPMARVLLPYVHGQSSPELGPRLRQVRLQALYPPDPGPRVHTCSAAGFSTSDYDV 3783  
QY 476 -----H-----Y-----L-----NG--Q-----VL-----L 484  
Db 3784 GWESPHNGSGTWAYSAPDLGAWSGSCAVYDGGYVQELGSLSESRDLRLFLQLHNWL 3843  
QY 485 -N-----L-----GR----- 488  
Db 3844 DNRSAVFLLETRYSPAVGLHAAVTLRLEFFPAAGRALAALSVRPFALRRLSAGLSPLLT 3903  
QY 489 SICL-----Y-----T-----L----- 495  
Db 3904 SVCLLLFAVFAVAEARTWHREGRVRLGAWARWLLVALTAALVRLAQIGAARQW 3963  
QY 496 -----P-----H-----H----- 498  
Db 3964 TRFVRGRPRFTSFQVAHVSSAARGLAASLLFLLLVKAQHVRFVRQWSVFGKTLCLRAL 4023  
QY 499 -----LVV-----I----- 502  
Db 4024 PELLGVTGLVGLVAYAQALAILLVSSVDLSVAQALLVLCPTGLSTLCPAESMHLS 4083  
QY 503 PL--V-----AL--I-----ELL-----V-----L 513  
Db 4084 PLLCGLMALRWGALRLGAVILRWYHARLGRYPAWEPQDYENVELFLRRLRLMGL 4143  
QY 514 -K 514

Db 4144 SK 4145

RESULT 14

US-09-479-467A-2

; Sequence 2, Application US/09479467A

; Patent No. 6723557

; GENERAL INFORMATION:

; APPLICANT: Sternberg, Paul W.

; APPLICANT: Barr, Maureen M.

; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE HOMOLOGS REQUIRED FOR MALE MAT

; TITLE OF INVENTION: BEHAVIOR IN NEMATODES AND ASSAYS BASED THEREON

; FILE REFERENCE: 18021-2901B

; CURRENT APPLICATION NUMBER: US/09/479,467A

; CURRENT FILING DATE: 2000-01-06

; PRIOR FILING DATE: 60/115,127

; PRIOR FILING DATE: 1999-01-06

; NUMBER OF SEQ ID NOS: 16

; SOFTWARE: Patentin Ver. 2.0

; SEQ ID NO 2

; LENGTH: 4303

; TYPE: PRT

; ORGANISM: Homo sapiens PKD-1 protein

US-09-479-467A-2

Query Match 69.3%; Score 2527.4; DB 4; Length 4303;

Best Local Similarity 10.0%; Pred. No. 1.5e-31;

Matches 410; Conservative 82; Mismatches 14; Indels 3576; Gaps 355;

Qy 1 MSIS-----S-----D-----E-----EVN-----F-----10

Db 72 LDVSHNLLRALDVGLLANLSALAELDLSNNKISTLEBEGIFANLNLSEINLSGNPFECDC 131

Qy 11 -LVY-----R-----Y-----15

Db 132 GLAWLPQWABEQVRVVPPEAATCAGPGSLAGPILLGILLDSGCGEYVACLPDNSSGT 191

Qy 16 -----LQ-----E-----SG-20

Db 192 VAAVFSAAHEGLLOPACSAFCFTGQGLAALSEQWCLCGAASPSSAFACLSLCSGP 251

Qy 21 -----F--S-----H--S-----AF-----T--FG--29

Db 252 PAPAPTCRGPTLLQHVFPASPGATLVGPHGLASGLAAFAHIAAPLPVTDTRWDFDGS 311

Qy 30 --I-----KSH-----I-----S-QSN-38

Db 312 AEVDAAGFAASHRYVLPGRYHTAVLALGAGSALLGTDVQVEAAPAALELVCPSSVQSD 371

Qy 39 -----I-N-----G-----AL-----V-----P-----P-----46

Db 372 SLDSLQNRGSGLEAAYSIVALGEEPARAVHPLCFSDTEIFPONGHCYRLVVEKAALQ 431

Qy 47 -----AA-----LIS-----I-----IQ-K-----G-----L--57

Db 432 AQEQCQAWAGAAWVDSPAVQRFVSRVTRSLDVMIGFSTVQGVGVPAPQGEAFSLES 491

Qy 58 -Q-----EDG-----YV-B-----AE-----V-----S--65

Db 492 CONWLPGEHPATAEHCVRLGPTGWCNTDLCASPHSYVCELPQCGPVQDAENLLVGAPSG 551

Qy 66 -----IN-----EDG-----TL--F-----73

Db 552 DLQGPLTFLAQDGLSAPHEVPMVFPGLRLSREAFITAEFGTQELRRPAQLRLQVYR 611

Qy 74 -----D-----GR--P-----I-----78

Db 612 LLSTAGTPENGSEPSRSPNRTQLAPACMPGKRCPCGANICLPDASCHPQACANGCTS 671

Qy 79 -----E--S-----LSLI-----DAV-M-P-DVV-----92

Db 672 GPGLPGAPYALWREFLFSVPAGPPAQYS-VTLHGQD-VLMLPGDLVLQHDAGFGALLHC 729

Qy 93 -----Q--T-----R--Q--Q-----97

Db 730 SPAPGHPGPRAPYLSANASSWLPPLPAQLEGTCGCPACALRLLAQREQTLVLLGLRPNPG 789

Qy 98 -----A-----Y-----RD-KL-----A--Q--105

Db 790 LRLPGRYEVRAEAVGNGVSRHNLSCSPDVVPVAGLVYIPAPDPGRGLYVPTNGSALVLQV 849

Qy 106 -----QQ-----A--A-----A-----A-----112

Db 850 DSGANATATARWPGGSLSFENVCVPALVATFVPACPFMETNDTFSVWALPMLSEGEHV 909

Qy 113 -----A--A-----A-----A-----A-----118

Db 910 DVVENSASRANLSLRVTAEEPIGRLRATPSPEARVLQGVLYRSPVVEAGSDMFRWTI 969

Qy 119 -----S-----Q-----Q-----S--A--124

Db 970 NDKQSLTFQNVVFNVIYQSAAVFKLSLTASNHVSNVTNVNVTVERMNRMQGLQVSTVPA 1029

Qy 125 -----KN-----GE-----N-----TA-----131

Db 1030 VLSPNATLALTAGVLVDSAVEAFLWTFDGEQALHQFPYPYNESFPVDPSPVAQLVEH 1089

Qy 132 N-----GE-----EN-----G--A-----H--139

Db 1090 NVTHYTAAPGEYLLTVLASNAFENLTQQVPVSRASLPVAVGVSDGLVAGRPVTFYPH 1149

Qy 140 -----T-----I-----ANNHT-----146

Db 1150 PLPSPGGVLYTDFDGGSPVLTSQSPAAANTVARSCTYHVRLEVNNTVSGAAAQADVRF 1209

Qy 147 -----DM-M--E-----V-----D-GD-----154

Db 1210 BELRGLSVDMSLAVEQAPVWVSAAVQTGDNITWTFDMDGTGLSGPEATVEHVYLAQN 1269

Qy 155 -----VE-----IP--P-----NKA-----162

Db 1270 CTVTVGAGSPAGHARSLLHVLVFLVLEVEPAACIPTODARLTAYVTGNPAHYLFDWT 1329

Qy 163 -----V-----V--L-R-G-----H-----ESEV-----172

Db 1330 FGDGSSNTTVRGCPVTTHNFTSRGTFPLALVLSRRVRAHYFTSICVEPEVGNVTLQPER 1389

Qy 173 -FI-----CAW-----NP-VS-----D--LL--AS--186

Db 1390 QFVQLGDEAWLVACAMPFPFPRYTYWDFGTEEAAPTRARGPEVTFIYRDPGSYLVTVTASN 1449

Qy 187 -----GS-----GD-----190

Db 1450 NISAANDSALVEQEPVLVTSIKVNGSLGLELQQLPFLSAVGRGPASVYLDLGDGWL 1509

Qy 191 -----S-----TARI--WN-L-SE-----N-----S--202

Db 1510 GPEVTHAYNSTGDFTVRVAGMNEVSERSEAWNLVTKRRVRLVWNASRTVPLNGSVSFS 1569

Qy 203 TS--GS-TQL--VL--RHC-----I-----RE-G-----G-QD-----221

Db 1570 TSLEAGSDVRYSNVLCDR-CTPIPGPTISYTFRSVGTNIIIVTAENVEGSAQDSIFVV 1628

Qy 222 -----V-----PSN-----KD--V--T-----SL--231

Db 1629 LQLIEGLQVVGGRYFPNTNHTVQLQAVVRDGTNVSYSWTAWDRDGPALAGSGKFSLTVL 1688

Qy 232 -----D-----M-----N-----234

Db 1689 EAGTVHQLRATNMLGSAWADCTMDFEVPGWMLVAASPNPAVNTSVTLSELAGSGV 1748

Qy 235 -----S--EG-----T--L-LAT-----GS-----YD-----246

Db 1749 VYTWLSEEGSWETSEPFTHSPPTPGLHLVMTAGNPLGASNATVEVDVQVPSGLSIR 1808

Qy 247 -----G-F--A-----R-----I--W-----TK--D-G-----NL--258

Db 1809 ASPPGGSFVAAGSVFPWGLATGTVNSWCAVPGSSKRGPHVTWVFPDAGTFSIRLN- 1867  
QY 259 AS- -T- -L- -GO- -H- -K- -G- -G- -267  
Db 1868 ASNAWSVSTYNLTAEPIVGLVWASSKVAPQQLVHFQIILLAGASAVTFLQVGGAN 1927  
QY 268 P-I- -P- -AL- -K- -W- -N- -KKG- -278  
Db 1928 PEVLPGFRFSGHSPRGDHHVSVRGKXHVSWAQVRIIVLEAVSGLOVPCCEPGIATG 1987  
QY 279 -NP- -ILS- -AG- -285  
Db 1988 TERNFTARVQGRSVAYAFYSIQVQGDLSVLISGRDVTYTEVAAGLLEIQVAFNALG 2047  
QY 286 -V-D- -K- -T- -T- -II- -WD- -294  
Db 2048 SENRTLVEQDVAQVVALQSGPCFTNRSQAFAATSPSRVAYHWDGDSGGPGQDTDE 2107  
QY 295 -A-HT- -G- -E- -A- -KQ- -Q- -FP- -305  
Db 2108 PRAEHSVLRPGDVRVQVNASNLVSFFVAQATVTVQVLACREPEVDVVLPLQVLMRBSQRN 2167  
QY 306 F- -H- -SA- -P- -AL- -DV- -314  
Db 2168 YLEAHVLDLRCVITYOTEYRWEVYRTASQCPGRPARVALPGVDVSRPRLVPLRALPVGH 2227  
QY 315 -D- -D- -W- -W- -QS- -N- -319  
Db 2228 YCFVVFVSFGDTPLTOSIQANVTVAERLVPPIIEGGSYRVMSDRLVLGDSESYDPNLE 2287  
QY 320 -N- -TF- -322  
Db 2288 DGDQTPLSFHWCAVSTQREAGCALNFGPRGSSVTIPIERLAGVEYTFSLTVWKAGR 2347  
QY 323 -A- -SC- -325  
Db 2348 KEEATNQTVLIRGRPIVLSVCCKAQAVYEVRSVYVLEGRCLNCSGSKRGWAA 2407  
QY 326 -S- -T- -D- -M- -C- -I- -331  
Db 2408 RTFSNKLVLDETTTSGAGMRLVLRGVLRDGEGYTTTLTVLGRSGBEGCASIRLSP 2467  
QY 332 -H- -V- -C- -334  
Db 2468 NRPLPGSCRLPFLGAVHALTTKVHECTGCHDAEDAGAPLVYALLRRCRQCHCEFCV 2527  
QY 335 -K- -LG- -337  
Db 2528 YKGSLSYGAVLPPGPRPHEVGLAVVQDLGAAVVVALNRSLAITLPEPNGSATGLTVW 2587  
QY 338 -Q-D- -R- -P- -341  
Db 2588 LHGLTASVLPGLLRQADPOHVIEYSALVTVLNEYERALDVAABPKHERQHAQIRKNIT 2647  
QY 342 -I- -K-T- -FO- -346  
Db 2648 ETLVSLRVTVDDIQIAAALACQMGPSRELVCRLCKQTLHKLEMMMLILQAEATTAGTV 2707  
QY 347 -G- -H- -T- -349  
Db 2708 TPTAIGDSLITNITGDLIHLASSDVRAPOPSSELGAESPMSRWASQAYNLTSALMILMRSR 2767  
QY 350 -NE- -351  
Db 2768 VLNEEPLTLAGEIIVAQKRSRPSRLCYGGAPGCHFSIPEAFSGALANLSDVVQLIF 2827  
QY 352 -V- -N- -AI- -K- -W- -357  
Db 2828 LVDSNPPFGYISNYTVSTKVASMAFQOAGAQIPIERLASERAITVKVPNNSDMAARGH 2887  
QY 358 -D- -P- -TG- -N- -LLA- -365

Db 2888 RSSANSANVVVQPOASVAVVTLSSNPAAGLHLQNLNTLLDGHYLSBEEPEPYLAVYLH 2947  
QY 366 S- -CS- -D- -D- -370  
Db 2948 SPPRNEHNCASRRIRPESLOGADHRPYTFFISPGSRDPAGSYHLNLSHFMSALQVS 3007  
QY 371 -M-T- -L- -KIM- -S- -MK- -Q- -380  
Db 3008 VGLYTSLCQYFSEEDVMWRTGELLPLEETSQPAVCLTRHLTAFGASLFPVPPSHVRVFP 3067  
QY 381 -D-N- -C- -V- -H- -DLQ- -QNN- -K- -EYTIK 398  
Db 3068 EPTADVNYIMLTCAVCLVTYVMMAAILHKLD-QLDASGRAIPFCGQRGRFYEL-VK 3125  
QY 399 -W- -S- -PT- -G- -P- -GT- -406  
Db 3126 TWGRGSGTTHVGMLYGVDSRSHRHLDGDRAFHRNSLDIFRIATPHSLGSSWKIRVW 3185  
QY 407 -NN- -P- -N- -AN- -L- -M- -L- -AS- -419  
Db 3186 HDNKGSLPAWFLQHVIVRDLQTARSAPFLVNDWLSVETEANGGLVEKEVLAASDAALLRF 3245  
QY 420 -FD- -S- -T- -V- -R- -L- -W- -D- -428  
Db 3246 RRLVAELQRFKHIWLSIWDPPRSRFTRIQATCCVLLICLFLGANAVMYGAVGDS 3305  
QY 429 -VD- -R- -G- -432  
Db 3306 AYSTGHVSRSLPSLSDTVAVGLVSSVVVYVYLAILFLPRMSRKVAGSPSPAGQOVL 3365  
QY 433 -I- -C- -H- -TLT- -439  
Db 3366 DIDSLDSSVLDSFLTFTSGLHAEQAFVQOMKSDLPDDSKSLVCWPSGEGTILSWPDLLS 3425  
QY 440 -K- -440  
Db 3426 DPSIVGSLNLQARGOAGHGLGPEEDGFLSPSPAKSPKSFASDEDLIOQLAEGVSSPA 3485  
QY 441 -H- -Q- -E- -443  
Db 3486 PTQDTHMETDLSSSSTPGEXTETLALQRLGELPPSPGLNWQEQQAARLSRTGLVEGL 3545  
QY 444 -P- -V- -Y- -SVA- -449  
Db 3546 RKRLLPANWASLHGLSLLLVAVAVSGVGVGASFPFGVSVAVMLSSSSASFLASFLGWEP 3605  
QY 450 -FS- -PD- -453  
Db 3606 LKVLLEALYFSLVAKRLHPDEDDTLVESPAVTPVSGARVPRVPPHGFALFLAKEEARKVK 3665  
QY 454 -G- -R- -Y- -LAS- -GSFD- -KC- -H- -I- -468  
Db 3666 RUHGMRLSLLVYMLFLVTLTASYG--DASCHGHAYRLQSAIKQELHSAFLAITRSEEL 3723  
QY 469 -W- -N- -T- -OV- -C- -L- -475  
Db 3724 WFWMAHVLLPYVHGQSSPELQPPRLRQVRLQEAALYDPDPGPRVHTCSAAGSFSTSDYDV 3783  
QY 476 -H- -Y- -L- -NG- -Q- -VL- -L- -484  
Db 3784 GWESPHNGSGTWAYSAPDLLLGAWSGCAVYDGGVQELGSLSEBSRDLRPLQLHNWL 3843  
QY 485 -N- -L- -GR- -488  
Db 3844 DNRSAVLELTRYSPAVGLHAAVTLRUEFFAAGRALAALSVRPFALRRLSAGLSLPLIT 3903  
QY 489 SICL- -Y- -T- -L- -495  
Db 3904 SVCLLLFAVHFAVEARTWHREGVRVRLGAWARLLVALTAATALVRLAQIQAADROW 3963  
QY 496 -P- -H- -H- -498  
Db 3964 TRFVRGRPRRFTSFQVAHVSSAARGLAASLLFLLLVKAAQHVRFVRQWSVFGKTLCLRAL 4023

QY 499 -----LVV-----I-----502  
Db 4024 PELGVTGLVGLVAVYQAAILLVSSCVDSLWSVAQALLVLCGTGLTCLPAESWHL 4083  
QY 503 PL--V-----AL-----I-----ELL-----Y-----L 513  
Db 4084 PLLCVGLWALRWLWGLRGLGAVILRWRYHALRGELYRPANWEPQDYEMVELFRLRLWML 4143  
QY 514 -K 514  
Db 4144 SK 4145

## RESULT 15

US-09-052-469-8  
; Sequence 8, Application US/09052469  
; Patent No. 6380360  
; GENERAL INFORMATION:  
; APPLICANT: Harris et al.  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: One Financial Center  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk, 3.50 inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/052,469  
; FILING DATE: Concurrently herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/422,582  
; FILING DATE: 14-April-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9507766.5  
; FILING DATE: 13-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9411900.5  
; FILING DATE: 14-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB PCT/GB94/02822  
; FILING DATE: 23-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9326470.3  
; FILING DATE: 24-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams, Ph.D., Kathleen M.  
; REGISTRATION NUMBER: 34,380  
; REFERENCE/DOCKET NUMBER: 3265/74165  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 345-9100  
; TELEFAX: (617) 345-9111  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4302 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-052-469-8

Query Match 69.2%; Score 2526.5; DB 3; Length 4302;  
Best Local Similarity 9.9%; Pred.No.1.5e-31;  
Matches 405; Conservative 13; Mismatches 13; Indels 3575; Gaps 351;  
QY 1 MSIS-----S-----D-----EVN-----F-----10

Db 72 LDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEGIPANLFNLSEINLSGNPPECDC 131  
QY 11 -LVY-----R-----Y-----15  
Db 132 GLAWLPRWAEQQVRVVPPEAATCAGPGSLAGQPLGIPLLDSGCCGEEYVACLPDSSGT 191  
QY 16 -----IQ-----E-----SG- 20  
Db 192 VAAVSFAAHEGLLOPEACSAFCSTGQGLAALSEQWCLCGAAQPSSASFACLSICSGP 251  
QY 21 -----F-S-----H-S-----AF-----T-FG-----29  
Db 252 PPPAPTCTGPTLLQHVFPASPATLVGPHGPLASQGLAAFAHPAALPVTATRWDFDGS 311  
QY 30 --I-----KSH-----I-----S-OSN- 38  
Db 312 AEVDAAGPAASHRYVLPGRYHVTAVALGAGSALLGTQVVEAAPAAALELVCPSSVQSD 371  
QY 39 -----I-N-----G-----AL-----V-----P-----46  
Db 372 SLDSLTONRGSGLEAAYSIVALGEEPARAVHPPLCPSTDTEIFPGNGHCYRLVVEKAWLQ 431  
QY 47 -----AA-----LIS-----I-----IQ--K-----G-----L-- 57  
Db 432 AQEQCOAWAGALAMVDSPAVQRFVLSRVTRSLDVWIGFSTVQGVVEGAPQGEAFSLES 491  
QY 58 -Q-----YV-E-----AE-----63  
Db 492 CQNLPGEPHATAEHCVRGLGPTGMCNTDLCAPHSYVCELOPGGVPQDAENLLVGAPSG 551  
QY 64 -----V-----S-----I-----66  
Db 552 DLQGLTPLAQDGLSAPHEPVEVFPGLRLSREAFLTABFGTQELRRPQLRLQVYR 611  
QY 67 -----N-----E-----D-----G-T- 71  
Db 612 LLSTACTPENGSEPESSPDNRTQLAPACMPGRCWCPGANICLPDASCHPQACANGCTIS 671  
QY 72 -----LF-----D-----74  
Db 672 GFLPGAYALWREFLFSVAAGPPAQSVTLHGQDVLMLPGDLVGLQHDAGPGLLHFCSP 731  
QY 75 -----G-RP-----77  
Db 732 APCHPGQAPYLSANASSWLPPLPAQLEGTWACPCALRLLAATEQLTVLLGLRPNPGLR 791  
QY 78 -----I-E-----S-----L-I-----D- 85  
Db 792 MGRYEVRAEVGNGVSRHNLSCSFDVSPVAGLRVIYPAPRDGRLVYPTNGSALVLQVDS 851  
QY 86 -----A-----V-MP-----DV 91  
Db 852 GANATATARPQGSVSARFENPCPALVATFVPGCPMETNDTLFVVVALPWLSEGEHVVDV 911  
QY 92 V-----O-T-----R-----O-Q-----A-Y-----R-----D 101  
Db 912 VVENSASRANLSLRVTAEPIGLRATPSPEARVLQGVLRYSPPVVEAGSDMVFRWTIND 971  
QY 102 KLAQ-----Q-----QAAAA-----AAA-----AS-----Q-Q-----121  
Db 972 K--QSLTFQNVVFNVIYQSAAVFKLSLTASNHSNVTVNVTVERMNRMQGLQVSTVPA 1029  
QY 122 -----G-----SA-K-----N-----GE-----N-----TA-----131  
Db 1030 VLSPNATLVLTGGLVDSAVEFAFLWNFGDGEQALHQFQPPYNESFPVDPDSVAQVLVEH 1089  
QY 132 N-----GE-----EN-----G-----A-----H 139  
Db 1090 NVNHTYAAEGEYLLTTLASNAFENLTQQVPVSVRASLPVAVGVSDGLVAGRPVTFYPH 1149  
QY 140 -----T-----I-----ANNHT-----146

Db 1150 PLSPGGLVLTWDFGDSVLTQSPAANHTYASRGTYHVRLVNNVTSGAAAQADRVRF 1209  
QY 147 ---DM-M-E---V---D-GD---154  
Db 1210 EELRGLSDMSLAVEQAPVVVSAAVTQDNITWTFDMGDTVLSPGEATVEHVYLAQN 1269  
QY 155 ---VE---IP-P---NKA---162  
Db 1270 CTVTGAASPAHLARSLHVLVFLVLEVRPEAPACIPTQDARLTAVVTGNPAHYLFDWT 1329  
QY 163 ---V---L-R-G---H---ESV---172  
Db 1330 FGDSSNTVRCPTVTHNFRSGTFPLALVLSRVNRAHYFTSICVEPEVGNVTLOPER 1389  
QY 173 -FI---CAW---NP-VS---D-LL---AS-186  
Db 1390 QFVLGDEAWLVACANPPPPYRYTWDGTEAAPTRARGEVTFIYRDPGSYLVTASN 1449  
QY 187 ---GS---GD---190  
Db 1450 NISAANDSALVEQEPVLVTSIKVNGSLGLELOQPYLFSAVGRGPASLYLDLGDGWL 1509  
QY 191 ---S---TARI-WN-L-SE---N---S 202  
Db 1510 GPEVTHAYNSTGDFTRVAGMNEVRSEAWLNVTKRRVGLVNASRTVVPLNGSVSFS 1569  
QY 203 TS---GS-TOL-VL-RHC---I---RE-G---G-OD---221  
Db 1570 TSLEAGSDVRYSVLCDR-CTPIPGGTISYTRSVGTENIIYTAENEVGSAGDSIFVV 1628  
QY 222 ---V---PSN---KD-V---T---SL---231  
Db 1629 LQLEGLQVVGGRYFPTNHTVQLQAVRDGTNVSYSWTAWRDRGPALAGSGKFSITVL 1688  
QY 232 ---D---W---N---234  
Db 1689 EAGTYHVQLRATNMLGSAWADCTMDFEVPVGLMTVTASNPAAVNTSVTSLAELAGSGV 1748  
QY 235 ---S-EG---T-L-LAT---GS---YD---246  
Db 1749 VYTWSLEGLSWETSBPFTTHSPPTGLHLVTWTAGNPLGSANATVEVDVQVPVSLSR 1808  
QY 247 ---G-F-A---R---I---W---TK---D-G---NL 258  
Db 1809 ASEPGGSFVAAGSSVPFWGQLATGTNVSMCAVPGGSKRGPHVTWVPDAGTFSIRLN 1867  
QY 259 AS---T---L---QO-H---K---G---267  
Db 1868 ASNAVSWSATYNLTAEPIVGLVLWASSKVAPGQLVHFQIILAAGSAVTFRLQVGGAN 1927  
QY 268 P-I---F-AL---K---W---N-KKG---278  
Db 1928 PEVLGPRFHSPPRVGDHVVSVRGKNHVSMAQVRIVLEAVSLQMPNCCPEGIATG 1987  
QY 279 ---NF---ILS---AG---285  
Db 1988 TERNFTARVQGRVAVAYFSLQKQDLSVLISGRDVTYTPVAAGLLEIQVRAFNALG 2047  
QY 286 ---V-D---K-T---T---II-WD---294  
Db 2048 SENRTLVLVEQDAVQVVALQSGPCFTNRSQAPEAATSPRPRVAYHWDGDSGPGQDTDE 2107  
QY 295 ---A-HT---G-E-A---KO---Q---FP---305  
Db 2108 PRAHSYLRPGDVRVQVNASNLVSFFVAQATVTVQLACREPEVDVVLPLQVLMRMSQRN 2167  
QY 306 F---H---SA---P-AL---DV---314  
Db 2168 YLEAHVDLRDCTVYQTYEYRWEVVTASQCPGPRPARVALPGVDVSRPRLVPLRALPVGH 2227  
QY 315 ---D---W---QS---N---319  
Db 2228 YCFVFWVSGDPTLTQSIQANTVAPERLVPPIEGGSYRWSDTRDLVLDGSESYDPNLE 2287

QY 320 ---N---TF---322  
Db 2288 DGDQTPLSFWHACVASTQREAGCALNFGPRGSSTVTIIPRERLAAGVEYTFSLTVMWKA 2347  
QY 323 ---A---SC---325  
Db 2348 KEBATQTVLIRSGRPVIVSLECVSCKAQAVYEVSRSSVYVLEGRCLNCSGSKRGRWAA 2407  
QY 326 ---S-T---D---M---C-I---331  
Db 2408 RTEFNKTLVLDETSTTSAGMRLVLRGVLRDGEGYTFTLTVLGRSGEBEGCASIRLSP 2467  
QY 332 ---H---V---C-334  
Db 2468 NRPLGSGCELPLGAVHALTKVHECTGWHDAEDAGAPLVYALLLRRCRQCHCEFCV 2527  
QY 335 -K---LG---337  
Db 2528 YKGLSSYGAVALPPGPRPHEVGLAVVQDQCGAAVVALNRSIAITLPEPNSATGLTVW 2587  
QY 338 ---Q-D---R---P---341  
Db 2588 LHGLTASVLPGLLRQADPOHVIEYSIALVTVLNEYERALDVAABPKHERQHRQAIRNIT 2647  
QY 342 ---I---K-T---FO---346  
Db 2648 ETIVSLRVHTVDDIOQIAAALACQMPSELVCRSCCLKQTLHKLLEAMLLILOAETAGTV 2707  
QY 347 ---G-H---T---349  
Db 2708 TPTAIGDSILNITGDLIHLASSDVRAPOPSSELGAESPRMVASQAVNLSALMRLMSR 2767  
QY 350 ---NE---351  
Db 2768 VLNEEPLTLAGEBEIVAGKRSDFRSLCYGAPGCHFSIPEAFSGALANLSDVVQLIF 2827  
QY 352 -V---N---AI-K---W---357  
Db 2828 LVDSNPPFGYISNYTVSTKVASMAFOTQAGAQIPIERLASERAITVKVPNNSDMAARGH 2887  
QY 358 ---D-P-TG---N---LLA---365  
Db 2888 RSSANSANVVQPQASVGAVTLDSSNPAAGHLQLNVTLLDGHYLSBEPEPYLAVYLH 2947  
QY 366 S---CS---D---D---370  
Db 2948 SEPRNEHNCASRRIRPESLQCADHRPYTFFISPGSRDPAGSYHLNLSHFRWSALQVS 3007  
QY 371 ---M-T-L---KIW-S---MK---Q---380  
Db 3008 VGLYTSLCQYFSEEDMVRTEGLLPLEETSPPQAVCLTRHLTAFGASLFPVPPSHVRFPVP 3067  
QY 381 ---D-N---C---V---H-DLQ---QHN-K-EIYTIK 398  
Db 3068 EPTADVNYIVMLTCAVCLVTVYVMAAILHKLD-QLDASGRAIPFCQGRGRFKYEIL-VK 3125  
QY 399 ---W---S-PT---G---P---GT---406  
Db 3126 TGMRGSGTTAHVGMILYGVDSRSGHRHLDGDRAFHRNSLDIFRIATPHSLGSMVKIRVW 3185  
QY 407 -NN---P---N---AN-L-M-L-AS-AS-419  
Db 3186 HDNKGSLPANFLOHVIVRDLQTAARSFAFFLVNDWLVSVEETANGELVKEVLAASDAALLRF 3245  
QY 420 ---FD---S-T-V-R---L---W---D-428  
Db 3246 RRLVVAELQRFQKHILWLSIWDPRPSRFTRIQATCCVLLJCLFLGANAVYVAGVDS 3305  
QY 429 ---VD---R---G---432  
Db 3306 AYSTGHVSRLSPLSVDTVAUGLVSSVVYVYLAITLFLPMSRSKVAGSPSPTPAGQVQL 3365

QY 433 -I--C---I-----H-----TLT----- 439  
Db 3366 DIDCLDSSVLDSSFLTFSLHAEAFVGQMKSDLFLDDSKSLVCWPFSGEGTTLSPDLLSD 3425  
QY 440 -----K----- 440  
Db 3426 PSIVGNLRLQARGHGLGPEEDGFSLASPYSPAKSFASDEDLIQOVLAEVGSPPAP 3485  
QY 441 ---H-----O-----E--- 443  
Db 3486 TQDTHMETDLLSSSLSTPGKTTETLALQRLGELGPPSPGLNWEQQAARLSRTGLVEGLR 3545  
QY 444 ---P-----V--Y-----SVA----- 449  
Db 3546 KRLLPAWCA SLANGLSLLVAVAVAGWVGASFPPGVSVAWLLSSSASFSLASFLGWEPL 3605  
QY 450 -----FS-----PD----- 453  
Db 3606 KVLLEALYFSLVAKRLHPDDEDTLVESPAVTPVSARVPRVPPHGFALFLAKEARKVKR 3665  
QY 454 --G--R---Y-----LAS--GSPD-KC-----V---H---I----- 468  
Db 3666 LHGMLRSLLYMFLVLVTLLASYG--DASCHGAYRLQSAIKOELHSRAPLAITRSEELW 3723  
QY 469 -W-----N--T-----QV-----C-----L--- 475  
Db 3724 PNMAHVLLPYVHGNQSSPELGPRLQVRLQALYDPDPPGPRVHTCSAAGGFSTSDYDVG 3783  
QY 476 ---H-----Y-----L-----NG--Q-----VL-----L- 484  
Db 3784 WESPHNGSGTWAYSAPDLLGAWSGWCAVYDGGYVQELGSLSESRDLRFLQLHNWLD 3843  
QY 485 N-----L-----CR-----S 489  
Db 3844 NRSRAVLELTRYSPAVGLHAAVTLRLEPPAAGRALAALSVRPFALRLRLSAGLSPLITS 3903  
QY 490 ICL-----Y-----T-----L----- 495  
Db 3904 VCLLLFAVHPAVAEARTWHEGRWVRLGLGAWARWLLVALTAALVRLAQLGAADQWT 3963  
QY 496 -----P-----H-----H----- 498  
Db 3964 RFVGRPRRFTSFDQVAHVSSAARGLAASLLFLLLVKAAQHVRVFRQWSVFGKTLCLALP 4023  
QY 499 -----LVV-----I-----P 503  
Db 4024 ELLGVTLGLVVLGVAVAYQAAILLVSSCVDLSWVAQALLVLCPTGLSTLCPAESWHLSP 4083  
QY 504 L--V-----AL-----I-----ELL-----V-----L- 513  
Db 4084 LLCVGLWALRLWGLRLGAVILRWRYHALRGELYRPAWEPQDYEMVELFURRLRWGLS 4143  
QY 514 K 514  
Db 4144 K 4144





Db 172 ATAAATTSAGVSKQNSKREAVTNGEENPAHSV-NNHA-KPNEIDGVEIPSSKATVLR 229  
QY 167 GHESEVFIICAWNVPVSDLLASGSDSTARIWNLSNSTSGSTQVLRHRCIREGGQDVPSNK 226  
Db 230 GHESEVFIICAWNVPVSDLLASGSDSTARIWNLSNSTSGSTQVLRHRCIREGGHDPVPSNK 289  
QY 227 DVTSLDWNSGTTLATGSDYDGFARIWTKDGNLASTLGOHKGPFPALKWKNKGNFIILSAGV 286  
Db 290 DVTSLDWNTGTTLATGSDYDGFARIWTKDGNLASTLGOHKGPFPALKWKNKGNFIILSAGV 349  
QY 287 DKTITIIWAHTGAKEAQOFPFHSAPALDWDQSN--TFASCSSTDMCHVCKLGDRIKTF 345  
Db 350 DKTITIIWAHTGAKEAQOFPFHSAPALDWDQ--NNTTFASCSSTDMCHVCKLGDRIKTF 408  
QY 346 QGHTNEVNAIKWDPGTGNLLASCSDMTLTKIWSMKQNCVHDLOHNEKIYTIKWSPTGPG 405  
Db 409 QGHTNEVNAIKWDPGTGNLLASCSDMTLTKIWSMKQNCVHDLOHNEKIYTIKWSPTGPA 468  
QY 406 TNNPNANMLASFSFSTVRLWDVDRGICHTLTKHOEPVYSVAFSPDGRYLASGSPDKC 465  
Db 469 TSNPNANMLASFSFSTVRLWDVDRGICHTLTKHOEPVYSVAFSPDGRYLASGSPDKC 528  
QY 466 VHTWNTQ---VCLH-Y--LNG--QVLLN-----LGR-----SICLYTLPLHLVWIPLVA 506  
Db 529 VHTWNTQSGNLV--HSYRGTTGGIFEVCCWARGDKVGASGDSGVC-----V- 572  
QY 507 LIELLVL-K 514  
Db 573 L-D---LRK 577

RESULT 2  
US-10-264-049-2892  
; Sequence 2892, Application US/10264049  
; Publication No. US2004005579A1  
; GENERAL INFORMATION:  
; APPLICANT: Birse et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PAI33PI  
; CURRENT APPLICATION NUMBER: US/10/264,049  
; CURRENT FILING DATE: 2002-10-04  
; PRIOR APPLICATION NUMBER: PCT/US01/18569  
; PRIOR FILING DATE: 2001-06-07  
; PRIOR APPLICATION NUMBER: US 60/209,467  
; PRIOR FILING DATE: 2000-06-07  
; NUMBER OF SEQ ID NOS: 4360  
; SOFTWARE: PatentIn Ver. 3.1  
; SEQ ID NO 2892  
; LENGTH: 542  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (45)  
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
US-10-264-049-2892

Query Match 84.0%; Score 3064.6; DB 15; Length 542;  
Best Local Similarity 73.2%; Pred. No. 1.9e-43;  
Matches 416; Conservative 45; Mismatches 23; Indels 84; Gaps 27;

QY 5 SDEVNF---LV-----YR-----YLOESGFSHSAFTGKSHISQSNING 41  
Db 1 TDC--FKESKVDLQSQKPKRTHLTCGCTGTLTC-GFSHSAFTGKSHISQSNING 57  
QY 42 ALVPPAALISITQKGLQVVAEVSINEDGTLFDRPIESLSLIDAVNPVQVTRQAYRD 101  
Db 58 TLVPPAALISITQKGLQVVAEVSINEDGTLFDRPIESLSLIDAVNPVQVTRQAFRE 117  
QY 102 KLAQQQ--AAAAA--AAASQ-----QG-S-----AKNGTANGENGAAHTTANNHTDM 148  
Db 118 KLAQQQASAAAAAATAAATAATTTAGVSHQNPKNREATVNGEENRAHSV-NNHA-K 175

QY 149 -MEVDGVEIPPNKAVLRLGHESEVFIICAWNVPVSDLLASGSDSTARIWNLSNSTSGST 207  
Db 176 PMEIDGVEIPSSKATVLRHRCIRESEVFIICAWNVPVSDLLASGSDSTARIWNLSNSTSGST 235  
QY 208 QVLRHRCIREGGQDVPSNKDVTSLDWNSGTTLATGSDYDGFARIWTKDGNLASTLGOHKG 267  
Db 236 QVLRHRCIREGGHDPVPSNKDVTSLDWNTGTTLATGSDYDGFARIWTKDGNLASTLGOHKG 295  
QY 268 PIFALKWKNKGNFIILSAGVDKTTIIWAHTGAKEAQOFPFHSAPALDWDQSN--TFASCS 326  
Db 296 PIFALKWKNKGNFIILSAGVDKTTIIWAHTGAKEAQOFPFHSAPALDWDQ--NNTTFASCS 354  
QY 327 TDMCHIVCKLGDRIKTFQGHTEVNAIKWDPGTGNLLASCSDMTLTKIWSMKQNCVHD 386  
Db 355 TDMCHIVCKLGDRIKTFQGHTEVNAIKWDPGTGNLLASCSDMTLTKIWSMKQNCVHD 414  
QY 387 LQOHNEKIYTIKWSPTGPGTNNPNANMLASFSFSTVRLWDVDRGICHTLTKHOEPVY 446  
Db 415 LQAHNEKIYTIKWSPTGPGTNNPNANMLASFSFSTVRLWDVDRGICHTLTKHOEPVY 474  
QY 447 SVAFSPDGRYLASGSPDKC VHTWNTQ---VCLH-Y--LNG--QVLLN-----LGR----- 488  
Db 475 SVAFSPDGRYLASGSPDKC VHTWNTQSGNLV--HSYRGTTGGIFEVCCWARGDKVGASD 532  
QY 489 -SICLYTLPLHLVWIPLVA LIELLVL-K 514  
Db 533 GSVK-----V-L-D---LRK 542

RESULT 3  
US-10-363-829-446  
; Sequence 446, Application US/10363829  
; Publication No. US20040142331A1  
; GENERAL INFORMATION:  
; APPLICANT: Jackson, Stuart E.; Lincoln, Stephen E.;  
; APPLICANT: Altus, Christina M.; Dufour, Gerard E.;  
; APPLICANT: Chalup, Michael S.; Jackson, Jennifer L.;  
; APPLICANT: Jones, Anissa L.; Yu, Jimmy Y.;  
; APPLICANT: Wright, Rachel J.; Gietzen, Darryl;  
; APPLICANT: Liu, Tommy F.; Yap, Pierre E.;  
; APPLICANT: Dahl, Christopher R.; Momiya, Monika G.;  
; APPLICANT: Bradley, Diana L.; Rohatgi, Sameer D.;  
; APPLICANT: Harris, Bernard; Roseberry Lincoln, Ann M.;  
; APPLICANT: Gerstein, Jr., Edward H.; Peralta, Careyna H.;  
; APPLICANT: David, Marie H.; Panzer, Scott R.;  
; APPLICANT: Flores, Vincent Z.; Daffo, Abel;  
; APPLICANT: Marwaha, Rakesh; Chen, Alice J.;  
; APPLICANT: Chang, Simon C.; Au, Alan P.;  
; APPLICANT: Inman, Rebekah R.  
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT  
; FILE REFERENCE: PT-1183 USN  
; CURRENT APPLICATION NUMBER: US/10/363,829  
; CURRENT FILING DATE: 2003-03-05  
; PRIOR APPLICATION NUMBER: PCT/US01/27628  
; PRIOR FILING DATE: 2001-09-05  
; PRIOR APPLICATION NUMBER: US 60/229,751  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 60/229,749  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 60/229,750  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 60/229,747  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 60/229,748  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 60/230,583  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: US 60/230,517  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: US 60/230,610  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: US 60/230,597



Db 841 VSSISQRTGALFILNLWASDKGYICEAENOFKIQSETTVTVTGLVAPLIGISPVAN 900  
 QY 96 ---QQ---A---Y---R---D---XL--- 103  
 Db 901 VIEGQQLTLPCTLLAGNPPIPERWKNKSNAMLLONFYITVRSQSLHIERVQLQDGEYTC 960  
 QY 104 ---A---Q---Q---AA--- 109  
 Db 961 VASNAVGTNNKTSVVVHVHPTIHHGQOILSTIEGIPVTLPCCKASGNPKPSVIMSKKGL 1020  
 QY 110 ---AAA---A---A---A---A--- 118  
 Db 1021 ISTSSAKFSAGADGSLYVSPGSEGEYVCTATNTAGYAKRKVQTVTVRFRVFDQBG 1080  
 QY 119 -SQ---A---Q---Q---GS--- 123  
 Db 1081 LSQDKPEISVLAGEEVLTPCEVKSLLPPIITWAKETQLISPPSRHTFLPSGMMKITET 1140  
 QY 124 ---A---A---K---N--- 126  
 Db 1141 RTSDSGMYLCVATNIAGNVTAQVKNLVHVPKIQRGPKHLKVQVQORVDIPCNAGOTPLP 1200  
 QY 127 ---GE---N---T---A---N---G--- 133  
 Db 1201 VITWSKGSTMLVDGHHVSNPDGTLISIDQATPSDAGIYTCVATNIAGTDETEILHVQ 1260  
 QY 134 ---E---N---GA---H--- 139  
 Db 1261 PPTVEDLEPPYNTTFQERVANQRIEPPCPAKGTPKPTIKLHNGRELTGREPGISILEDG 1320  
 QY 140 T---IA---N---N---HT---DM--- 148  
 Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVKDKQVTVNVSVLLNQLTNL 1380  
 QY 149 ---M---E--- 150  
 Db 1381 FCEVEGTPSIIMYKDNVQVTSSTIQTVNGKILKLFRATPDAGRYSCAKAINIAGTS 1440  
 QY 151 ---VD---G---DV---E---I---P---PN 160  
 Db 1441 QKYFNIDVLVPPTTIIGTFNFEVSNVLRDVALEQVKGTPFPDIHFKDGKFLGDPN 1500  
 QY 161 ---K---K---A---V---VL 165  
 Db 1501 VELLDRGOVLHKNARNNDKRYOCTVSNAGKQAKDIKLTINYPPSIKGNVTTDISVL 1560  
 QY 166 ---RG--- 168  
 Db 1561 INSLIKLECETRGLPMPAITWYKDGQPIMSSQALYIDKGQYLHI PRAQVSDSATYTCV 1620  
 QY 169 ---E---S---E---VPI---C---A---W--- 177  
 Db 1621 ANVAGTAESPHVDVYVPPMIEGNLATPLNKQVVIASHLTLECKAAGNPSPIITWLDKG 1680  
 QY 178 ---N--- 182  
 Db 1681 PVKANDNIRIBAGGKKEIMSAQEI DRGOYTCVATSVAGEKEIKYEVVDVLPVPAIEGDE 1740  
 QY 183 ---LL---ASGS---G---D---S--- 191  
 Db 1741 TSYFIVMVNLLLELDCHVTGSPPTIMWLKDGQILIDERDGFKILLNGRKLVIQAQVSN 1800  
 QY 192 ---TA---R---I---W 196  
 Db 1801 GLYRCMAANTAGDHKKFEVTVHVPPTIKSSGLSERVVVVKYKPVALQCIANGIENPSITW 1860  
 QY 197 ---NL--- 199  
 Db 1861 LKDDQPVNTAQGNLKIQSGRVLQIAKTLLLEDAGRYTCVATNAAGETQHQHVLHVHPPS 1920  
 QY 200 -E---N---S---SGST--- 207  
 Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPTIWKDNRLLSGSTMTFLNRQGIIDIES 1980

QY 208 -Q---L---V---LR-H---C---I--- 215  
 Db 1981 AQISDAGIYKCVAINSAGATELFYSLQVHVAPSIGSNMNVAVVNNPVRLECEARGIPA 2040  
 QY 216 ---R---E---G---Q---D--- 221  
 Db 2041 PSLTWLKDGPSSVSNGLQVLSSGRILALTSAQISDTGRYTCVAVNAAGEKORDIDLVR 2100  
 QY 222 -VPSN---KD---VT---S-L- 231  
 Db 2101 YVPPNIMEGEQNVLSIAVELLCQSDAIPPTTLTWLKGHGLLKKPGLSISENRSLVK 2160  
 QY 232 ---D---W---N---S---EG---TLL---A 241  
 Db 2161 IEDAQVDTGRYTCATNVAGTEKNVNNIWPVNIGSGDELTLQTLVIEGNLSLICES 2220  
 QY 242 T---GS---Y-D---G---F---A--- 249  
 Db 2221 SGIPPPNLIWKKKSPVLTDSMGRVRLSGRQLQISIAEKSDAALYSCVASNAGTAKK 2280  
 QY 250 ---R---I---WKDG--- 256  
 Db 2281 EYNLQVYIRPTITNSGSHPTIIVTRGKSLSECEVOGIPPTVTWMDKGHLIKAKGVE 2340  
 QY 257 ---NL-AS---T---L---QOH--- 265  
 Db 2341 ILDEGHILQLKNIHVSDTGRYCVAVNVAGTMDKKYDLSVHAPPSIIGNHRSPENISVVE 2400  
 QY 266 K---G---P-I---FALK-W--- 274  
 Db 2401 KNSVSLTCEASGIPLPSITWF--KDGWPVLSNSVRILSGRMLRMQTTMEDAGQYTCV 2458  
 QY 275 ---N---K-K---GN--- 279  
 Db 2459 VRNAAAGEERKIFGLSVLPPHIVGENTLEDVKVKEKQSVTLTCTEPTGPNVPEITWHKQGO 2518  
 QY 280 ---F---I---L---S---AGV--- 286  
 Db 2519 PLQDEEAHHIISGRFLQITNVQVPHGTGRYTCCLASSPAGHKRSFSLNVFVSPPTIAGVGS 2578  
 QY 287 D---K---T---I---I-W--- 293  
 Db 2579 DGNPEDEVTLNSPTSLVCEAYSYPATITWPKDGTPLESNRNIRILPGGRTLQILNAOE 2638  
 QY 294 D---A---H---TG--- 300  
 Db 2639 DNAGRYSCVATNEAGEMIKHYEVKYVPIPIINKGDLWGLSPKEVKIKVNNLTLECEA 2698  
 QY 301 ---K---Q---Q---F 304  
 Db 2699 YAIPTSASLSWYKDGQPLKSDDHVNIAANGHTLIQKEAQISDTGRYTCVASNIAGEDELDF 2758  
 QY 305 ---P-F---H--- 307  
 Db 2759 DVNIQVPPSFOKLWEIGNMLDTRNGEAKDVIINNPIISLYCETNAAPPPTLTWYKDGHP 2818  
 QY 308 -SA---P---A---LDV---D--- 315  
 Db 2819 TSSDKVLLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLPVPIIKGANS 2878  
 QY 316 ---WQ---SN--- 319  
 Db 2879 PEBVTVLVNSKALIECLSSGPAPRNSWQKQGPLEDHDKHFLSNRILQILNTOITDI 2938  
 QY 320 ---WT---FAS-C---S--- 326  
 Db 2939 GRVYCVAVNAGSAGAKYFNLVHVPVSPVIGPKSENLTVVVNNPISLTCEVSGFPPLSLW 2998  
 QY 327 ---T---D---MCI--- 331  
 Db 2999 LKNEQPIKLTNTLIVPGGRTLQIIRAKVSDGGEYTCIAINQAGEKSKKFLSVYVPPSI 3058

QY 332 --H-----V--C-----K-----L--GQ----- 338  
Db 3059 KDHDSLSVVNREGTSVLECSNAVPPVITWYKNGRMISTHVEILADQMLHIK 3118  
QY 339 -----D-----R-----P-IK-----TF----- 345  
Db 3119 KAEVSDTGQVCRAINVAGRDDKRNFLNVVPPSIEGPEREVIVETISNPVTLTCDATGI 3178  
QY 346 -----Q-----G-----H-----T----- 349  
Db 3179 PPPTIAWLKXHKRIENSLSLEVRILSGGSKLQIARSQHSNGNYTCIASNMEKAQKYF 3238  
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
Db 3239 LSIQPPSVAGABIPSDSVLLGENVELVCNANGIPTPLIQLWKDKPIASGETERIRVS 3298  
QY 361 --G--NLL-A--S-----C--S-----D-----D-----MTL----- 373  
Db 3299 ANGSTLNIYGALTSDTGKYTCVATNPAGEEDRIFNLNVVYPTIRGNKDEAEKMLTVD 3358  
QY 374 --K-----I-W-----SM-----KO--D--D-----NCV----- 384  
Db 3359 SINIECATGTPPPQINLWLNKGLPLPLSSHIRLLAAGQVIRVRAQVSDVAVVTCVASNR 3418  
QY 385 -----H-----Q-----D-----L--Q----- 388  
Db 3419 AGVDNKHYNLQVPAPPMNDMSMGTEBITVLKGSSTSMACITDGTAPAPMAWRDGPGLG 3478  
QY 389 -----Q-----K-----W-----H-N-----KEI 394  
Db 3479 DAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFILKVLPPHPSIEHEI 3538  
QY 395 -----YT-----I--K--W-----N----- 399  
Db 3539 SVIVNNPLETCTIASGIPAPKMTWKDGRPLPOTDQVQLGGEVLRISTAQVEDTGRYT 3598  
QY 400 -----SPTG-----P--GT-----N----- 408  
Db 3599 CLASSPAGDDKEYLVRVHPVPPNIAGTDEPRITVLRNQVTLCKSDAVPPPVITWLRN 3658  
QY 409 -----P-----N-----ANLM-LAS--A--S--F----- 420  
Db 3659 GERLOATPRVRLSGGRYLQINNADLGDYANTYCVASNIAGTKTREFILTVNVPNIKGG 3718  
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433  
Db 3719 POSLVILLNKSTVLECIAGGVPTPRITWRKOGAVLAGNHARYSILENGFLHIQSAHVDT 3778  
QY 434 -----C-----I--H-----TL-----T----- 439  
Db 3779 GRYLCWATNAAGTDRRRIDLQVHVPPSIAPGPTNMTVIVNVQTLTACEATGIPKPSINWR 3838  
QY 440 K--H-----Q-E-----P-V--Y-----S 447  
Db 3839 KNGHLNVQDNQNSYRLSSGSLVIPSVDATYECTVTNGAGDDKRTVDLTQVPPS 3898  
QY 448 VA-----F-S-----P--D----- 453  
Db 3899 IADEPTDLVTKHAPAVITCTAGVPFPPSIHWTNGIRLLPRDGYRILSSGAIEILATQ 3958  
QY 454 -----GRY-----D-----KCV----- 466  
Db 3959 LNHAGRYTCVARNAAGSAHRHVTLVHPEVPVQPQPSSELVHLNLPILLPCEATGTPSPF 4018  
QY 463 -----H-----I-W-----NT-Q--V 473  
Db 4079 PVIISPHLKEYIADVDPKPTILSCDAGLPPDPITWHKDGRAIVESIQRVLSGSLQIARV 4138  
QY 474 -----C-----L-----HY----- 477

Db 4139 QPGDAGHYTCMAANVAGSSSTSKLTAVHVPPIRSTEGHYTVNENSOAILPCVADGIPTP 4198  
QY 478 -----L-N-----GQ-----VL-----L-----N----- 485  
Db 4199 AINWKDNVLLANLLKGYTABPYGELILENVVLEDSGFYTCVANNAAAGETHHTVSLTVHV 4258  
QY 486 -----L--G--R-SIC-----L-YT-----LP-H-----H--LV----- 500  
Db 4259 LPTFTLPGDVLNKGELRLS-CKATGIPLPKLTWTFFNNIIPAHFDSVNGHSELVIER 4317  
QY 501 -----V--I-----P-----L-----V-----A-L----- 507  
Db 4318 VSKEDSGTYVCTAENSVGFKAGFYVYKPEPPVFKGDYPSNMWIEPLGGNAILNCEVKGDP 4377  
QY 508 -----IE-----L-----L-VL-----K 514  
Db 4378 TPTIQNRKGVDLIEISHRIRQLGNGSLAIYGTVNE 4412

RESULT 5  
US-10-032-189-128  
; Sequence 128, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkete, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenda  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/279,840  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/282,981  
; PRIOR FILING DATE: 2001-04-11  
; PRIOR APPLICATION NUMBER: 60/283,656  
; PRIOR FILING DATE: 2001-04-13  
; PRIOR APPLICATION NUMBER: 60/309,247

; PRIOR FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: 60/311,754
; PRIOR FILING DATE: 2001-08-17
; PRIOR APPLICATION NUMBER: 60/313,331
; PRIOR FILING DATE: 2001-08-17
; NUMBER OF SEQ ID NOS: 260
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 128
; LENGTH: 5636
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3003)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3041)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3367)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; US-10-032-189-128

Query Match
Best Local Similarity 71.2%; Score 2596.5; DB 14; Length 5636;
Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

QY 1 M-S-----I-----SS--D-----EV-----8
Db 1 MISWEVHTVFLFALLYSSLAQDASQSEIRAEFFEGASTLAFVFDVTGSMYDDLQVVI 60
QY 9 -----NF-IV-----12
Db 61 EGASKILETSKRPRKPLNFALPHFDHPGPIVTTIDPKPKFYELRELYVQGGDCPE 120
QY 13 -----Y-----R--Y-----I-Q-----17
Db 121 MSGAIAKIALEISLPGSFIVFTDARSKOYRLTHEVLQLIQKQSQVVFVLTGDCDDRTH 180
QY 18 -----E-----SG--F-----S-H-S-----A-----F 26
Db 181 IGYKVEELASTSGGVFHLDDKKQVNEVLKWEAEVQASKVHLLSTDHLEQAVNTWRIPF 240
QY 27 -----T-----I-----FG-----I-----K 31
Db 241 DPLKEVTVSLSGPSPMIEIRNPLGLIKKGFGLHELLNIHNSAKVNVNKEPEAGMWTX 300
QY 32 -----S--H-----I-----S--O-----36
Db 301 TSSGRHSVRITGLSTIDFRAGFSRKTLDPKKTVPVQGIPTYYVLLNTSGISTPARID 360
QY 37 -----S-----S-----NI-----N-----G-----41
Db 361 LLELLSISGSLKTIIPVKKYPRKPYGIWNIISDFVPPNEAFFLKVTKYDKDDYLFQRVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
Db 421 VFSFISVPDAPKVTMPKTPGYLQPGQIPCSVDLSLLPFTLSFVRNGVTLGVDQYLKESA 480
QY 48 -----A--L-----I-----S-----IIQ-----KG-----56
Db 481 SVSLDIKVTLSDEGFYECIAVSSAGTCRAQTFDFVSEPPVPIQVPPNVTTPGERAVLT 540
QY 57 -----L--Q-----Y-----V--E-----61
Db 541 CLIIISAVDYNLTQNRDRVLAEPARITRLANLSLELKSFKFNDAGEYHCHMSSEGSS 600
QY 62 -A-----EVSI-----NE-----68
Db 601 AASVFLTVQEPKVTVMPPKQNSFTGGSEVSMCSATGPKPKIAWTNVDMEFVCSHYRM 660

QY 69 --DGLTF-----D-----G-----R-----P-----I-- 78
Db 661 TSDGLFIKNAAPKADAGIYGCLANSAGTDKNSFLRYIEAPKLMVVOSELLVALGDITV 720
QY 79 -E--S-----L--S--LI-----DA-----V-----87
Db 721 MECKTSGIPPPQVQKFGDLRLPSTFLIIDPLGLLKIQETQDLADAGDYTCVAINEAGR 780
QY 88 -----M-----P-DV-----VO-----T-----94
Db 781 ATGKITLDVGSPPVFIQEPADVMEIGSNVTLPCVVOGYVPEPTIKWRRLDNMPISRPFS 840
QY 95 -----R-----95
Db 841 VSSISQLRTGALFILNLNASDKGTIYCEAENQFGKIQSETTVTGTGLVAPLIGISPSVAN 900
QY 96 --QO-----A-----Y-----R-----D--KL-- 103
Db 901 VIEGQQLTLPCTLLAGNPPIERRWIKNSAMLLIQNFYITVRSDGLHIERVQLQDGGEYTC 960
QY 104 -----A-----Q--QO-----AA-----109
Db 961 VASNVAGTNNKTSVVVHVLPITIQHGQQLLSTIEGIPVTLPCKASGNPKPSVMSKKGL 1020
QY 110 ---AAA--A-A-----A--AA-A-----118
Db 1021 ISTSSAKPSAGADGSLVYVSPGESBESGYVCTATNTAGYAKRKVQVTVYVRPRVFGDLRG 1080
QY 119 -SQ-----Q-----GS-----123
Db 1081 LSQDKPVEISVLAGEBVTLPCEVKSLLPPIITWAKETQLISFPSPRHTLPSSGSMKITET 1140
QY 124 -----A-----K-----N-----126
Db 1141 R1SDSGMYLCVATNAGNTQAVKLVNHPVKIQRGPKHLKVQVQGVQVDPICNAQGTPLP 1200
QY 127 -----GS--N-----T--A-N-G-----133
Db 1201 VITWSKGGSTMLVDGEBHVSNDPDTLSIDQATSPDAGIYTCVATNIAGTDETEITLHVQE 1260
QY 134 -----E--N-----GA-----H-----139
Db 1261 PPTVEDLEPPYNTTFOERVANORIEPPCPAKGTPKPTIKWLHNGRELTCOREPGISILESG 1320
QY 140 T--IA-----N-----N--HT-----DM-----148
Db 1321 TLLVIASTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVKDKEQVSNVSVLLNQLTNL 1380
QY 149 -----M-----E-----150
Db 1381 FCEVEGTPSPIIMWYKDNVQVTESTIQTNNNGKILKLFRATPDADAGRYSCKAINIAGTS 1440
QY 151 -----VD-----G-----DV--E-----I-----P-----PN 160
Db 1441 QKYFNIDVLVPTIITGNFPKESVSVLNDRDVALEQCQVKGTPFPDIIHFKDGGKFLFLGDPN 1500
QY 161 -----K-A-----V-----VL 165
Db 1501 VELLDRGVHLKVNARRNDKGRYQCTVSNAAAGQAKDIKLIYIIPPSIKGNVTDTISVL 1560
QY 166 -----RG-----C-----H-----168
Db 1561 INSLIKLECKTRGLPMPAITWYKDGQFIMSSSQALYIDKQYLIHPRAQVSDSATYTCVH 1620
QY 169 -----E-S-----E-----VF1-----C--A-----W-----177
Db 1621 ANVAGTAESFFHVVDVVPVPMIEGNLATPLNKQVIAHSLTLECAAGNPSPILTWLKDG 1680
QY 178 -----N-----P-V--SD-- 182
Db 1681 PVKANDNFRIEAGGKKLEIMSAQIEDRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740
QY 183 -----LL-----ASGS-----G-----D-----S-- 191

Db 1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKQGLIDDERDGFKILLNGRKLIVIAQAVSNT 1800  
QY 192 -----TA-----R-----I-W 196  
Db 1801 GLYRMAANTAGHKKEFEVTVHVPPTIKSSGLSERVVVKYKPVALQCIANGIPNPSITW 1860  
QY 197 -----NL-----S 199  
Db 1861 LKDDQPWNTAQGNLKIQSSGRVLQIAKTLLLEDAGRYTCVATNAAGETQOHIQLHVHEPPS 1920  
QY 200 -E-----N-----S-----SGST----- 207  
Db 1921 LEDAGKMLNETVLVSNPVLQCKAAGNPVPIITWYKDNCLLSGSTSTWFLNRQOIIDIES 1980  
QY 208 -Q-----L-V-----LR-H-----C-----I-- 215  
Db 1981 AQISDAGIYKCAINSAGATELFYSIQVHVAPISGSNNMVAVVNNPVRLECEARGIPA 2040  
QY 216 -----R-----E-G-----G-Q--D-- 221  
Db 2041 PSITWLKDGSPVSSPNSGLVLSGRILALTSTQISDTGRYTCVAVNAAGKORDIDLAV 2100  
QY 222 -VPSN-----XD-----VT--S-L- 231  
Db 2101 YVPPNIMGEEQNVLSIQAVELLQSDAIPPTLTWLDKGHPHLLKPKGLSISENRSVLK 2160  
QY 232 -----D-----W--N--S-----EG--TLL-- 240  
Db 2161 IEDAQVODGRYTCENTVAGTEKKNNVNIHVPNIGSDELTQLTWIEGNLISLCE 2220  
QY 241 AT-----GS--Y-D-----G-----F-----A- 249  
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVRIILSGGQLOISAEKSDAALYSCVASNAVAGTAK 2280  
QY 250 -----R-----I-----WTKDG----- 256  
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSISLECEVQGIPTVTWTKDGHPLIKAKGV 2340  
QY 257 -----NL-AS-----T-----L-----GQH----- 265  
Db 2341 EILDEGHILQKNIHSDTRYCVVAVNAVAGMTDKKYDLSVHAPPSIIGNHRSPEINISV 2400  
QY 266 -K-----G-PI--FALK--W----- 274  
Db 2401 EKNSVSLTCEASGIPLPSTTW--KDGWPSVLSNVRILSGGRLMLQTMEDACQYTC 2458  
QY 275 -----N-----K-K-----GN----- 279  
Db 2459 VVRNAAEERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPBITWHKDG 2518  
QY 280 -----F-I-----L-----S--AGV- 286  
Db 2519 QPLQDEAHIIISGRFLQITNVQVPHTRYTCCLASSPAGKRSFSLNVFVSPTIAGV 2578  
QY 287 -D--K--T-----T--I-----I-W----- 293  
Db 2579 SDGNPEDVTILNSPTSLVCEAYSYPATITWFKDGTPLESNNRILPGRTLOILNAQ 2638  
QY 294 -D-----A-----H-----TG-----E 299  
Db 2639 EDNAGRYSCVATNAEAGEMIKHEVKVYIPPIINKGDLWGLSPKEVKIKVNTLTLECE 2698  
QY 300 A-----K-----Q----- 303  
Db 2699 AVAIPASISLWYKQGPLKSDDHVNTAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758  
QY 304 F-----P-F-----H- 307  
Db 2759 FDNVQVPPSFQKLWEIGNMLDTRNGEAKDVIIINNPISLYCETNAAPPTTWTYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315

Db 2819 LTSSDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYVRVLVPPPIEGANS 2878  
QY 316 -----WC-----SN----- 319  
Db 2879 LPEEVTVLVNSKALIECLSSGSPAPRNSWQKQOQPLEDDHHKFLNGRILQILNTOITD 2938  
QY 320 -----NT-----FAS--C-----S 326  
Db 2939 IGRYVCVAENTAGSACKYFNLVHVPVPSVIGPKSENLTVVVNNFISLTCEVSGFPDPDL 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKNXQPIKLNNTLIVPOGRTLQIIRAKVSDGGEYTCIAINXAGESKKFSLTVVYPPS 3058  
QY 332 --H-----V--C-----K-----L--GO-- 338  
Db 3059 IKDHDSELSVNVNREGTSVLECESNAVPPPVITWYKGRMITESTHVEILADQOMLHI 3118  
QY 339 -----D-----R-----P-IK----- 343  
Db 3119 KKAESVDTGOYVCRAINVAGRODDKNFHLNVVYPPSIEGPEREVIVETISNPVTLTCDATG 3178  
QY 344 --T-----F-----Q--G-----H-----T----- 349  
Db 3179 IPPPTIAMLKNYKRIENSLSLEVRILSGGSKLQIARSOHSDSGNYTCIASNMEGKAKY 3238  
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
Db 3239 FUSIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIQMLKGKPIASGETERIRV 3298  
QY 361 --G--NLL-A--S-----C--S-----D-----D-----MTL-- 373  
Db 3299 SANGSTLNIYGALTSDTGKYTCAVNPAGEEDRIFNLNVVYVPTIRGNKDEAEKLMYVD 3358  
QY 374 --K-----I-W-----SM-----KQ--D-----NCV-- 384  
Db 3359 TSINIECRXTGTPPOINWLKNGLPPLSSHIRLLAAGOVIRIVRAQVSDVAVYTCASN 3418  
QY 385 -----H-----D-----L--Q-- 388  
Db 3419 RAGVDKNHYNQVAFAPPNMDNSMGTEETIVLKGSSSTSMACITDGTAPSPAMWLURDQOPLG 3478  
QY 389 -----Q-----H-N-----KE 393  
Db 3479 LDAHLLTVSTHGMVQLLKAETDSGKYTCIASNEAGEVSKHFLKVLPEPHINGSEHEE 3538  
QY 394 I-----YT-----I--K--W----- 399  
Db 3539 ISVIVNNPLELTCIASGIPAPKMTWMDGRPLPQTDQVTLGGGEVLRISTAQVEDTRY 3598  
QY 400 -----SPTG-----P--GT-----N----- 407  
Db 3599 TCLASSPAGDDDDKEYLVRVHVPPIAGTDEPRDITVLRNRQVLTCECKSDAVPPPVITWLR 3658  
QY 408 N-----P-----N-----ANLM-LAS--A--S--F----- 420  
Db 3659 NGERLOATPRVILSGRYLOINNADLGTANTYTCVASNIAGTKTREFILTVNVPPNIK 3718  
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433  
Db 3719 GPQSLVILLINKSTVLECIAEAGVPTPRIWTKDGAVLGNHARYSILENGFLHIQSAHVT 3778  
QY 434 -----C-----I--H-----TL-----T----- 439  
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHVPPIAGPTNMTVINVQTLTACBATGIPKPSINW 3838  
QY 440 -K--H-----Q--E-----P-V--Y----- 446  
Db 3839 RKNHLLNVQDNQNSRYLLSSGLVILSPSDDTATYECTVTNGAGDDKRTVDLTVOVPP 3898  
QY 447 SVA-----F-S-----P--D----- 453  
Db 3899 SIADPTDFLTKHAPAVITCTASGVPPPSIHWTKNGIRLLPRGDCYRILSSGAIBLAT 3958

QY 454 -----GRY-----L-----ASG--S- 461  
DB 3959 QLNHAGRYTCVARNAAGSAHRHVTLVHBPVPIQOPSELHVLNPPILLPCEATGTPSP 4018  
QY 462 F-----D-----KCV----- 466  
DB 4019 FITWQEGINVTSGRNVHVLPSGGLQISRAVPEDAGTYMCVAQNPAQTALGKIKLVQV 4078  
QY 467 -----H-----I-W-----NT-Q----- 472  
DB 4079 PPVISPHLKEYVIADVDPITLSCADGLPPDITWHKDGRAIVESIRQVLSGSLQIAF 4138  
QY 473 V-----C-----L-----HY----- 477  
DB 4139 VQPGDAGHYTCMAANVAGSSSTKLTUVHPPRIRSTEGHYTNWNSQAILPCVADGIPT 4198  
QY 478 -----L-N-----GO-----VL-----L-----N----- 485  
DB 4199 PAIWKKNVLLNLLGKYTAEPYGGELLENVLEDSGFTCVANNAAGEDTHVSLTVH 4258  
QY 486 -----L-G-----R-SIC-----L-YT-----LP-H-----H-LV----- 500  
DB 4259 VLPTFTLPGDVSINKGEQLRLS-CRATGIPLPKLTWTNNIIPAHFDSVNGHSELVIE 4317  
QY 501 -----V-----V-----I-PL----- 505  
DB 4318 RVSKBDSGTYVCTAENSVGFKATGFVYVKEBPVFKGDYPSNWIEPLGNAILNCEVKGD 4377  
QY 506 ---AL-----IE-----L-----L-VL-----K 514  
DB 4378 PTFTIQNRKGVDEISHRIRQLNGSLAIYGTVNE 4413

RESULT 6

US-10-120-801-72  
; Sequence 72, Application US/10120801  
; Publication No. US20030203843A1  
; GENERAL INFORMATION:  
; APPLICANT: Pena, Carol  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Kekuda, Rameesh  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Menraban, Fuad  
; APPLICANT: Topper, James N.  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Wasserman, Scott  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Gunther, Erik  
; APPLICANT: Komuves, Laszlo  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-340  
; CURRENT APPLICATION NUMBER: US/10/120,801  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/285748  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 60/286068  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/286292  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/288334  
; PRIOR FILING DATE: 2001-05-03  
; PRIOR APPLICATION NUMBER: 60/291241  
; PRIOR FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: 60/322284  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/285609  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: Patent in Ver. 2.1

; SEQ ID NO 72  
; LENGTH: 5636  
; TYPE: PRT  
; ORGANISM: human  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (3003)  
; OTHER INFORMATION: Wherein Xaa is any amino acid.  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (3041)  
; OTHER INFORMATION: Wherein Xaa is any amino acid.  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: (3367)  
; OTHER INFORMATION: Wherein Xaa is any amino acid.  
; US-10-120-801-72  
  
Query Match 71.2%; Score 2596.5; DB 14; Length 5636;  
Best Local Similarity 9.7%; Pred. No. 2.7e-29;  
Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;  
  
QY 1 M-S-----I-----SS--D-----EV----- 8  
DB 1 MISWEVHTVFLFALLYSLAQDASPQSIRAEPEFGASTLAFVFDVTGSMYDDLQVI 60  
QY 9 -----NF-LV----- 12  
DB 61 EGASKILETSLKRPKRPLENFALVPHDPEIGFVITTDPKKFQYELRELYVQGGDCPE 120  
QY 13 -----Y-----R-----Y-----LQ----- 17  
DB 121 MSIGAIIKIALEISLPGSFYIVFTDARKDYRLTHEVLQIQKQSQVFLTGDCDDRTH 180  
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26  
DB 181 IGYKVEEIASTSGQVFLDKKQVNEVLKVEEAQVSKVHLLSTDHLEQAVNTWRIFP 240  
QY 27 -----T-----F-----I-----K 31  
DB 241 DPSLKEVTVLSGSPSEIIRNPLGKIKKGFLHLLHNSAKVNVVKEPEAGMTVK 300  
QY 32 ---S-H-----I-----S--Q----- 36  
DB 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFKKTVSRRPVQGIPTYVLLNTSGISTPARID 360  
QY 37 -----S-----N-----G----- 41  
DB 361 LLELLSISGSLKTIPTVYVPHRKYGIWNIISDFVPNEAFKLVGTGDKDYLFORVSS 420  
QY 42 -----A-----L-----VP-----P-----A 47  
DB 421 VSFSSIVDPAPKVTMEKTPGYVLPQGPQPCSVDSLLPFTLSFVRNGVTGLGVQYKLKESA 480  
QY 48 -----A-----L-----I-----S--IIQ-----KG----- 56  
DB 481 SVSLDIKAVTLSDGEFYECIAVSSAGTGRAQTFFVDVSEPPVIVQVPPNVTVTTPGERAVLT 540  
QY 57 -----L--Q-----Y-----V--E----- 61  
DB 541 CLIIISAVDYNLTWQRNDRDLVLAEPARITLANLSLELSKVKFNDAEYHCWVSSGGSS 600  
QY 62 -A-----EVSII-----NE----- 68  
DB 601 AASVFLTVQEPKVTVMKQNSFTGGSEVSIIMCSATGYPKPKIAWTVNDMFIVGSHRYRM 660  
QY 69 --DGTLP-----D-----G-----R-----P-----I-- 78  
DB 661 TSDGTLFIKNAAPKDAIGYCLASNSAGTDKQNSTLRYTEAPKLMVQSELLVALGDTV 720  
QY 79 -E--S-----L--S--LI-----DA-----V----- 87  
DB 721 MECKTSGIPPPQVWFKGDLRLPSTFLIIDPLGLLKIQETQDLADGDYTCVAINAAGR 780



QY 88 -----M-----P-DV-----VQ-----T-----94  
Db 781 ATGKITLDVGPSPVFIQEPADVSGMEIGSNVTLPCYQGYPEPTIKWRLLDNMFISRPFS 840  
QY 95 -----R-----95  
Db 841 VSGISQRTGALFILNLWASDKGTICYEABNOFGKIQSETTVTVTVGLVAPLIGISPSVAN 900  
QY 96 -----QQ-----A-----Y-----R-----D--KL--103  
Db 901 VIEGQQLTLPCTLLAGNPPIPERWIKNSAMLLQNPIYTVRSOGSLHIERVOLQDGGEYTC 960  
QY 104 -----A-----O--QO-----AA-----109  
Db 961 VASNVAGTNNKTTVVVHVHPTTIHQOQILSTIEGIPVTLPCKASGNPKPSVIWSKKGEL 1020  
QY 110 -----AAA--A-A-----A--AA--A-----118  
Db 1021 ISTSAKFAGADGSLYVVSPEGESEGYVCTATNTAGYAKRKVQLTVYVRPFGDLRG 1080  
QY 119 -SQ-----O-----Q-----GS-----123  
Db 1081 LSQDKVEISVLAGEEVLPCVEKSLPPPIITWAKETQLISPPSPRHTFLPSCMKITET 1140  
QY 124 -----A-----A-----K-----N-----126  
Db 1141 RTSDSGMVLCVATNIAGNVTAQVKNLVHVPKIQRGPKHLKVQVGQORVDIPCAOQTPLP 1200  
QY 127 -----GE--N-----T--A-N-G-----133  
Db 1201 VITWSKGGSTMLVDGHHVSNPDGTLISIDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260  
QY 134 -----E-----E--N-----GA-----H-----139  
Db 1261 PPTVEDLEPPYNTTFOERVANORIEPPCPAKGTPKPTIKWLHNGRELGTGREGISLEBG 1320  
QY 140 T--IA-----N-----N--HT--DM-----148  
Db 1321 TLLVIASVTPDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKEQVSNVLLNQLTNL 1380  
QY 149 -----M-----E-----150  
Db 1381 FCEVEGTPSPIIMWYKDNVQVTTESSTIQTVNNGKILKLFATPEDAGRYSCRAINAGTS 1440  
QY 151 -----VD-----G-----DV--E-----I-----P-----PN 160  
Db 1441 QKYFNIDVLVPPTIIGTNFPKEVSVVNLNRDVALLECQVKGTTPPDHWFKDGKPLFLGDPN 1500  
QY 161 -----V-----K-A-----V-----VL 165  
Db 1501 VELLDRGOVLHLKNARNDKGRYQCTVSNAGKQAKDIKLIYIPPSIKGNNVTTDISVL 1560  
QY 166 -----RG-----H-----168  
Db 1561 INSLIKLECKTRGLPMPAITWYKDGOPIMSSSOALYIDKQVLIHPRAQVSDSATVCHV 1620  
QY 169 -----E-S-----E-----VFI-----C--A-----W-----177  
Db 1621 ANVAGTAKSFHVTVVPPMIEGNLATPLNKQVVIASHSLTLECNAGNPSILTWLKGV 1680  
QY 178 -----N-----P-V--SD--182  
Db 1681 PVKANDNFRIEAGKKLEIMSAQEI DRGOYICVATSVAGEKEIKYEVDVLVPPAIEGDE 1740  
QY 183 -----LL-----ASGS-----G-----D-----S--191  
Db 1741 TSVFIWVNNLLELDCHVTGSPPTIMLKGDLIDERDGFILLNGRKLVIQAQVSN 1800  
QY 192 -----TA-----R-----I-W 196  
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPTIKSSGLSERVVVVKYKFPVALQCIANGIPNPSITW 1860

QY 197 -----NL-----S 199  
Db 1861 LKDDQPVNTAQGNLKIQSSGRVLQIAKTLEDAGRYTCVATNAAGETQQHIQLHVHPPS 1920  
QY 200 -E-----N-----S-----T-----SGST-----207  
Db 1921 LEDAGKMLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLLSGSTMFLNRQIIDIES 1980  
QY 208 -O-----L--V-----LR-H-----C--I--215  
Db 1981 AQISDAGIYKVCAINAGATELFYSLOVHVAPSISGSNNMVAVVNNPVPLECEARGIPA 2040  
QY 216 -----R-----E--G-----G--O--D--221  
Db 2041 PSLTWLKDGPVSSFFSNGQLVLSGGRILALTSQISDTGRYTCVAVNAAGEKORDIDLRV 2100  
QY 222 -VPSN-----KO-----VT--S-L--231  
Db 2101 YVPPNIMGEEQNVSLISOAVELLQSDAIPPTTLTWLKGHPHLLKPKGLSISENSVLK 2160  
QY 232 -----D-----W--N--S-----EG--TLL--240  
Db 2161 IEDAQVODTGRYTCBATNVAGTEKKNVNNVWVPNIGSDELTOLTWIEGNLISLCE 2220  
QY 241 AT-----GS--Y-D-----G-----F--A--249  
Db 2221 SSGIPPPNLIWKKKGPVLTDSMGRVRIILSGGRQLQISIAEKSDAALYSCVASNVAGTAK 2280  
QY 250 -----R-----I-----WTXDG-----256  
Db 2281 KEYNLQVYIRPTITNSGSHPTBIIVTRGKSI SLECEVOGIPPTVTVWMDGHPHLLKAKGV 2340  
QY 257 -----NL-AS-----T--L--GOH-----265  
Db 2341 EILDEGHILQLKNIHVSDTGRYVCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400  
QY 266 -K-----G-PI--FALK--W-----274  
Db 2401 EKNSVLTCEASGIPLPSTTW--KDGWPSVLSNSVRIILSGGRMLRLMQTTMEDAQYTC 2458  
QY 275 -----N-----K-K-----GN-----279  
Db 2459 VVRNAAGEERKIFGLSVLPPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPBEITWHKDG 2518  
QY 280 -----F-I-----L--S--AGV--286  
Db 2519 QPLQDEAHIIISGGRFLQITNVQVPHTRYTCCLASSPAGHKRSRPSLNVFVSPITAGVG 2578  
QY 287 -D--K--T-----T--I-----I-W-----293  
Db 2579 SDGNPEDVILNSPTSLVCEAYSPPATITWFKDGTPLSNRNRIILPGRFLQILNAQ 2638  
QY 294 -D-----A--H-----TG-----E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKYIPIIPINKGLMGPGLSPKEVKIKVNTLTLECE 2698  
QY 300 A-----K-----O--Q-----303  
Db 2699 AYAI PASLSWYKDGQPLKDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758  
QY 304 F-----P-F-----H--307  
Db 2759 FDNVTOVPPSFQKLWEI GNMMLDTGRNGEAKDVI INNPISLYCETNAAPPTLTWYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315  
Db 2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGSDSLQYDVLVPPPIEGANS 2878  
QY 316 -----TA-----WQ-----SN-----319  
Db 2879 LPEEVTVLNKSALIECLSSGSPAPRNSMQDGQPLEDDHKKFLSNGRIQLINTQITD 2938  
QY 320 -----NT-----FAS--C--S 326

Db 2939 IGRYVCAENTAGSAKYNFLNHHVPPSVIGPKSENLTVVNNFISLTCEVSGFPPDDLS 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKNXQPIKLTNTLIVPGRTLIQIRAKVSDGGEVTCIAINKAGBSKKKFSLTIVVPPS 3058  
QY 332 ---H---V---C---K-----L---GO--- 338  
Db 3059 IKDHDSELSVNVNREGTSVLECESNAVPPVITWYKNGRMITESTHVEILLADGQMLHI 3118  
QY 339 ---D---R-----P-IK----- 343  
Db 3119 KKAESDGTQVCRAINVAGRDKNFNLNHHVPPSVIEGPEREVIVTISNPVTLTCDATG 3178  
QY 344 ---T---F-----Q---G-----H-----T----- 349  
Db 3179 IPPPTIAWLKNYKRIENSLEVRILSGGSKLQIARSQSDSGNYTCIASNMKGAKQYV 3238  
QY 350 -----N-E-V-NA-----IKW-D-P-----T----- 360  
Db 3239 FLSIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIOWLKDKGFIASGETERIRV 3298  
QY 361 ---G---NLL-A-S---C-S---D-----D-----D-----MTL-- 373  
Db 3299 SANGSTINIYGALTSDTGKTKVATNPAGEEDRIFNLNHHVPPSVTIRGNKDEAKLMTYVD 3358  
QY 374 ---K-----I-W-----SM-----KO-D-D-----NCV--- 384  
Db 3359 TSINIECRXTGTPPPQINLKNGLPLPSLSHIRLAAGOVIRVRAQVSDVAVYTCVASN 3418  
QY 385 ---H-----Q-----D-----L---Q--- 388  
Db 3419 RAGVDNKHYNLQVAPPNNMDSMGTEITVLKGSSTMACITDGTAPSNWALRDQOPLG 3478  
QY 389 ---L---VTHSGWVQLLKAEEDSGKYTCIASNEAGEVSKHFLKVLPEPHNGSEHEE 3538  
Db 3479 LDAHLTVSHGWLQLLKAEEDSGKYTCIASNEAGEVSKHFLKVLPEPHNGSEHEE 3538  
QY 394 I-----YT-----I-K-W----- 399  
Db 3539 ISVIVNNPLELTCTASGIPAPKMTWMDGRPLQTDQVOTLGGGEVLRISTAQVEDTGRY 3598  
QY 400 -----SPTG-----P---GT-----N----- 407  
Db 3599 TCLASSPAGDDKEXLVRVHVPNIAGTDEPRDITVLRNQVTLCKSDAVPPPVITWLR 3658  
QY 408 N-----P-----N-----ANLM-LAS-A---S-F----- 420  
Db 3659 NGERLOATPRVILSGRYLQINNADLGDYANTCVASNIAGKTREFILTVNVPNIKG 3718  
QY 421 -----DSTV-----RL-W-D-V-----DR-----G-I----- 433  
Db 3719 GPOSVLILNKSTVLECIAREGVPTPRITWRKQGAVLGNHARYSILENGFLHIQSAHVT 3778  
QY 434 ---C-----I---H-----TL-----T----- 439  
Db 3779 TGRYLCMATNAAGTDRRRIDLVHVPVSIAPGPTNMTVIVNQTTLACEATGPKPSINW 3838  
QY 440 -K-H---Q-E-----P-V-----Y----- 446  
Db 3839 RKNGHLLNVNDQNSYELLSSGSIIVISPSVDDTATVECTVNGAGDKETVLTQVVPF 3898  
QY 447 SVA-----F-S-----P-D----- 453  
Db 3899 SIADEPTDLVTKHAPAVITCTASGVFPFPIHMTKNGIRLLPRGDGVRILSSGAIBLAT 3958  
QY 454 ---GRY-----L---ASG--S- 461  
Db 3959 QLNHAGRYTCVARNAGSAHRHVTLVHVEPPVQOPQPSSEHLVILNPNLIPCEATGTPSP 4018  
QY 462 F-----D-----KCV----- 466

Db 4019 FITWQKEGINVNTSGRHNHVLPSGLQISRAVRVEDAGTYMCVAQNPAQTALGKIKLNVQV 4078  
QY 467 ---H---I-W-----NT-Q--- 472  
Db 4079 PPVISPFLKEYVIADVDPITLSCEADGLPPPDITWHKQGRAIVESIRQVLSGSLQIAF 4138  
QY 473 V---C-----L---HY----- 477  
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTVHVPPRIRSTEGHYTVNENSOAILPCVADGPT 4198  
QY 478 ---L-N-----GO---VL---L---N----- 485  
Db 4199 PAINWKONVLLANLLGKYTAEPYGEILLENVVLDSDSGFYTCVANNAGETHVTSLVTH 4258  
QY 486 ---L-G---R-SIC-----L-YT-----LP-H---H-LV-- 500  
Db 4259 VLPTFTFELPGLVSLNKGEOQLRLS-CKATGIPLPKLTWTNNIIPAHFDSVNGHSELVIE 4317  
QY 501 ---AL-----IE-----L---L-VL---K 514  
Db 4318 RVSKEDSGTYVCTAENSQVFKATGVYVYKPPVFKGDPSPNWIPLGNAILLNCEVKGD 4377  
QY 506 ---AL-----IE-----L---L-VL---K 514  
Db 4378 PTPTIQWNRKGVDDIETSHRIRQLGNGSLAIYGTVNE 4413

## RESULT 7

US-10-023-634-93  
; Sequence 93, Application US/10023634  
; Publication No. US20030236389A1  
; GENERAL INFORMATION:  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Colman, Steven D  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Ballinger, Robert A  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Shenoy, Suresh G  
; APPLICANT: Li, Li  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc  
; APPLICANT: Gusev, Vladimir Y  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Malyankar, Uriel M  
; APPLICANT: Gunther, Erik  
; APPLICANT: Smithson, Glenda  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Gerlach, Valerie  
; TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
; FILE REFERENCE: 21402-221  
; CURRENT APPLICATION NUMBER: US/10/023,634  
; CURRENT FILING DATE: 2002-06-28  
; PRIOR APPLICATION NUMBER: 60/256,025  
; PRIOR FILING DATE: 2000-12-15  
; PRIOR APPLICATION NUMBER: 60/265,163  
; PRIOR FILING DATE: 2001-01-30  
; PRIOR APPLICATION NUMBER: 60/272,929  
; PRIOR FILING DATE: 2001-03-02  
; PRIOR APPLICATION NUMBER: 60/274,864  
; PRIOR FILING DATE: 2001-03-09  
; PRIOR APPLICATION NUMBER: 60/276,688  
; PRIOR FILING DATE: 2001-03-16  
; PRIOR APPLICATION NUMBER: 60/277,880  
; PRIOR FILING DATE: 2001-03-22  
; PRIOR APPLICATION NUMBER: 60/286,409  
; PRIOR FILING DATE: 2001-04-25

;; PRIOR APPLICATION NUMBER: 60/309,246  
;; PRIOR FILING DATE: 2001-07-31  
;; PRIOR APPLICATION NUMBER: 60/315,600  
;; PRIOR FILING DATE: 2001-08-29  
;; NUMBER OF SEQ ID NOS: 132  
;; SOFTWARE: PatentIn ver. 2.1  
;; SEQ ID NO 93  
;; LENGTH: 5636  
;; TYPE: PR1  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: VARIANT  
;; LOCATION: (3003)  
;; OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the  
;; OTHER INFORMATION: specification.  
;; FEATURE:  
;; NAME/KEY: VARIANT  
;; LOCATION: (3367)  
;; OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the  
;; OTHER INFORMATION: specification.  
;; US-10-023-634-93

Query Match 71.2%; Score 2596.5; DB 14; Length 5636;  
Best Local Similarity 9.7%; Pred. No. 2.7e-29;  
Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

QY 1 M-S-----I-----SS--D-----EV-----8  
DB 1 MISWEVHTVFLFALLYSLAQAASQSEIRAEFFEGASTLAFVFDVTGSMYDVLQVI 60  
QY 9 -----NF-LV-----12  
DB 61 EGASKILETSKRPRPLFNALVPHDPPIGPVTTTDPKFOYELRELYVQGGDCPE 120  
QY 13 -----Y-----R--Y-----IQ-----17  
DB 121 MSIGAIALEISLPGSFYVFTDARSQYRLTHEVLQIQKQSQVVFVLTGDCDDRTH 180  
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26  
DB 181 IGKVVYEETASTSGGVFHLDKQVNEVLKWVEEAVQASKVHLLSTDHLEQAVNTWRIFF 240  
QY 27 -----T-----FG-----I-----K 31  
DB 241 DPSLKEVTVSLSGSPWIEIRNPLGLIKKGFGLHELLNIHNSAKVNVNKEPEAGMWTVK 300  
QY 32 -----S--H-----I-----S--Q-----36  
DB 301 TSSGRHVRITGLTIDFRAGFSRKPTLDFKKTVSRRVQGIPTYVLLNTSGISTPARID 360  
QY 37 -----S-----NI-----N-----G-----41  
DB 361 LLELLSISGSSSLTIPVKYYPHRKPGYIWNISDFVPPNEAFFLKVTGYDKDDYLFORVSS 420  
QY 42 -----A-----L-----VP-----P-----A 47  
DB 421 VSFSSIVDPKVTMEKTPGYVLPQGPICSVDSLLPFTLSFVRNGVTGLGVQYLKESA 480  
QY 48 -----A-----L-----I-----S-----IIQ-----KG-----56  
DB 481 SVSLDIKAVTLSDGEFYECIAVSAGTGFDFVSEPPPIQVNNVTVPGERAVLT 540  
QY 57 -----L--Q-----Y-----V--E-----61  
DB 541 CLIIISVDYNTWQRNDRVRLAEPARITLWLSLELKSVPFNDAAGEYCHWVSSEGGSS 600  
QY 62 -A-----EVSI-----NE-----68

DB 601 AASVFLTVQEPKPVTVMPKNQSFSGSEVSMCSATGYPKPKIAWTNDMFIVGSHRYRM 660  
QY 69 --DGLTF-----D-----G-----R-----P-----I--78  
DB 661 TSDGTLFIKNAAPKDAIGYCLASAGTDKQNSTURYIEAPKLMVMVQSELLVALGDITV 720  
QY 79 -E--S-----L--S--LI-----VO-----T-----94  
DB 721 MECKTSGIPPPQVQKPGDLELRPSTFLIDPLGLLKIQETQDLDAGDGYTCVAINAAGR 780  
QY 88 -----M-----P-DV-----V-----94  
DB 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRRLDNMPISRPFS 840  
QY 95 -----R-----95  
DB 841 VSSISQLRTGALFILNLWASDKGTICEAENQPKIQSETTVTVTGLVAPLIGISPSVAN 900  
QY 96 -----QQ-----A-----Y-----R-----D--KL--103  
DB 901 VIEGQQLTLPCTLLAGNPIPERRWIKNSAMLLQNPLYTVRSQGLHIERVQLQDGGEYTC 960  
QY 104 -----A-----Q--Q-----AA-----109  
DB 961 VASNAGTNNKTTVVVHVLPFTIQHQQILSTIEGIPVTLPCCKASGNPKPSVIWSKKGEL 1020  
QY 110 ---AAA---A-A-----A--AA--A-----118  
DB 1021 ITSSAKFAGADGSLYVVSPEGESEGYVCTATNTAGAKRVQLTVVYVPRVFGDLRG 1080  
QY 119 -SQ-----GE--N-----T--A--N--G-----123  
DB 1081 LSQDKPEISVLAGEEVLTPCEVKSLLPPIITWAKETQLISPPSPRHTFLPSCSMKITET 1140  
QY 124 -----A-----K-----N-----126  
DB 1141 RTSDSGMVLCVATNIAGNVQAVKLVNHPVKIQRGPKLVQVQGVQVDPICNAQGTPLP 1200  
QY 127 -----N-----T--A--N--G-----133  
DB 1201 VITWSKGSMTLVGDEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260  
QY 134 -----E--N-----GA-----H-----139  
DB 1261 PPTVEDLEPPYNTTFOERVANQRIEPPCPAKGTPKPTIKWLHNGRELTCREPGISILEEG 1320  
QY 140 T--TA-----N-----N--HT--DM-----148  
DB 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPPVVKDKQVSNVSVLLNQLTNL 1380  
QY 149 -----M-----E-----150  
DB 1381 FCEVEGTPSPIIMYKNDVQVTSSTIQTVNNKGILKLFATPDEADGRYSCKAINGTS 1440  
QY 151 -----VD-----G-----DV--E-----I-----P-----PN 160  
DB 1441 QKVFNIDVLVPTIIGTNPKEVSVVLRDVALEQCVKGTPTPDHIFWKGDKPLFLGDPN 1500  
QY 161 -----K-A-----V-----VL 165  
DB 1501 VELLDRGQVLHLKNARRNDKGRYQVCTVSNAGKQAKDIKLTIVIPPSIKGNVTTDISVL 1560  
QY 166 -----RG-----H-----168  
DB 1561 INSLIKLECKTRGLPMPAITWKGQPIMSSSQALYIDKGYLHPIRAQVSDSATYCHV 1620  
QY 169 -----E--S-----E-----VFI-----C--A-----W-----177  
DB 1621 ANVAGTAEKSFHVDVVPPIEMIEGNLATPLNKQVIAHSLTLECNAGNPSPILTWLDKGV 1680  
QY 178 -----N-----P--V--SD--182  
DB 1681 PVKANDNPFIEAGKKLEIMSAQEIQRGOYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740

QY 183 -----LL-----ASGS-----G-----D-----S-- 191  
DB 1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKQGLIDRDFKILLNGRKLVIQAQVSN 1800  
QY 192 -----TA-----R-----I-W 196  
DB 1801 GLYRCMAANTAGDKKEFEVTHVPTTIKSSGLSERVVVYKFPALQCIANGIPNPSITW 1860  
QY 197 -----NL-----S-----S 199  
DB 1861 LKDDQPNTAQNKLIOSSGRVLQIAKTLLEDAGRYTCVATNAAGETQHQIQLHVHEPPS 1920  
QY 200 -E-----N-----S-----SGST----- 207  
DB 1921 LEDAGKMLNETVLVSNPVLQCKAAGNPVPTIWKYKDNCLLGGSTSMWTFNLRGQIIDISS 1980  
QY 208 -Q-----L--V-----LR-H-----C-----I-- 215  
DB 1981 AQISDAGIYKCAINSAGATELFYSLOVHVAPSISGSNNMVAVVNNPVRLCEARGIPA 2040  
QY 216 -----R-----E-----G--Q--D----- 221  
DB 2041 PSLTWLKDGPVSSFSFNGVLQVLSGGRILALTSQISDTGRYTCVAVNAAGEKQORDILRV 2100  
QY 222 -VPSN-----KD-----VT--S-L- 231  
DB 2101 YVPPNMGEEQNVSVLISOAVELLQSDAIPPTTLTWLKDGHPLKPKGLSISENSVLK 2160  
QY 232 -----D-----W--N--S-----EG--TLL-- 240  
DB 2161 IEDAQVDTGRYTCBATNVAGKTEKQNVNINWPNIGSDDELTOITWIEGNLISLCE 2220  
QY 241 AT-----GS--Y--D-----G-----F-----A- 249  
DB 2221 SSGIPPPNLIWKKGSPVLTDSMGRVILSGGRQLQISIAEKSDAALXSCVASNAGTAK 2280  
QY 250 -----R-----I-----WTKDG----- 256  
DB 2281 KEYNLQVYIRPTITNSGSHPTIIVTRKGSISLCEVQGPPTVTWMDGHPLIKAKGV 2340  
QY 257 -----NL-AS-----T-----L-----GOH----- 265  
DB 2341 EILDEGHILQKNIHVSDTGRYCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPEINISVV 2400  
QY 266 -K-----G-PI-----FALK--W----- 274  
DB 2401 EKNSVSLTCEASGIPLPSTTF--KDGWPVSLNSVRILSGGRMLRLMOTTMEDAQYTC 2458  
QY 275 -----N-----K-K-----GN----- 279  
DB 2459 VVRNAGEERKIFGLSVLPVPHIVGENTLEBDVKVKEQSVTLTCEVTGNPVPBITWHKDG 2518  
QY 280 -----F--I-----I-W-----L--S--AGV-- 286  
DB 2519 OPLQDEAHIIISGGFLOITNVQVHTGRYTCCLASSPAGKSRSPSLNVFVSPTIAGVG 2578  
QY 287 -D--K--T-----T--I-----I----- 293  
DB 2579 SDGNPEDVTILNSPTSLVCEAYSYPATITWFKDGTPLSERNRILPGGRTLQILNAQ 2638  
QY 294 -D-----A-----H-----TG-----E 299  
DB 2639 EDNAGRYSCVATNEAGEMIKHYKVIPIINKDLMWGLSPKXVKIKVNNLTLECE 2698  
QY 300 A-----K-----Q-----Q----- 303  
DB 2699 AYAIPSASLSWYKDGQPLKDDHVNIAANGHTLIQKEAQISDTGRYTCVASNAGEDLD 2758  
QY 304 F-----P-F-----H----- 307  
DB 2759 FDNVIOVPPSFQKLWEIGNMLDTRNGEAKDVIINNPISLYCETNAAPPPTLTWYKDGHP 2818

QY 308 --SA-----P-----A-----LDV-----D 315  
DB 2819 LTSSDKVLLPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPIIEGANS 2878  
QY 316 -----WQ-----SN----- 319  
DB 2879 LPEEVTLVNKSALIECLSSGSPAPRNSWQDQPLEDDHFKFLSNGRILQILNTQITD 2938  
QY 320 -----NT-----FAS--C-----S 326  
DB 2939 IGRYVCVAENTAGSACKYFNLVNHPVPSVIGPKSENLTVVVNNFISLTCEVSGFPPDLS 2998  
QY 327 -----T-----D-----MCI----- 331  
DB 2999 WLKNYQPIKLTNTLIVPGRTRLQIIRAKVSDGGEYTCIAINXAGESKKKFSITVVVPPS 3058  
QY 332 --H-----V--C-----K-----L--GO-- 338  
DB 3059 IKDHDSESLSVVNVREGTSVLECESNAVPPVITWYKNGRMITESTHVEIILADGQMLHI 3118  
QY 339 -----D-----R-----P-IK----- 343  
DB 3119 KKAESDGTQYVCRAINVAGRDDKNFHLNVYVPPSIEGPEREVIVETISNPVTLTCDATG 3178  
QY 344 --T-----F-----Q--G-----H-----T----- 349  
DB 3179 IPPPTIAMLKNYKRIENSLSLEVRILSGSKLOIARSQHSDSGNYTCIASNMEGKAQYV 3238  
QY 350 -----G--NLL-A--S-----N-E-V-NA-----IKW--D--P-----T----- 360  
DB 3239 FLSIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIOWLKDGKPIASGETERIRV 3298  
QY 361 --G--NLL-A--S-----C--S-----D-----D-----MTL-- 373  
DB 3299 SANGSTLNIYGALTSDTGKYTCVATNPAGEDRIFNLNVYVPTTRIGNKDEAKMLTYVD 3358  
QY 374 --K-----I-W-----SM-----KQ--D--D-----NCV-- 384  
DB 3359 TSINIECRXTGTPPQINWLKNGPLPLPLSSHRLLAAGOVIRIVRAQVSDVAVYTCASN 3418  
QY 385 -----H-----D-----L--Q----- 388  
DB 3419 RAGVDKNHYLQVAFPPNNDMSMGTEETVLKGSSTSMACITDGTTPAPSMWLRDQPLG 3478  
QY 389 -----Q-----H-N-----KE 393  
DB 3479 LDAHLTVSTHGMVLQLLKAETDSGKYTCIASNEAGEVSKHFLKVLPEPPHNGSEHEE 3538  
QY 394 I-----YT-----I--K--W-----N----- 399  
DB 3539 ISVIVNNPLELTICIASGIPAPKMTWMDKGRPLQPTDQVTLGGGEVLRISTAQVEDTRY 3598  
QY 400 -----SPTG-----P--GT-----N----- 407  
DB 3599 TCLASSPAGDDKEVLRVHVPPNIAGTDEPRDITVLRNRQVTLCKSDAVPPPVITWLR 3658  
QY 408 N-----P-----N-----ANLM-LAS--A--S--F----- 420  
DB 3659 NGERLOATPRVILSGRYLQINNADLGDNTANTCVASNIAGKTTFRELFILTVNVPPNKG 3718  
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433  
DB 3719 GPQSLVILLNKSTVLECIAGEVPTPRIITWRKDGAVLGNHARYSILENGLFHQSAHVT 3778  
QY 434 --C-----I-----H-----TL-----T----- 439  
DB 3779 TGRYLCMATNAAGTDRRIDLQVHVPPSIAPGPTNMTVIVNVQTTILACEATGPKPSINW 3838  
QY 440 -K--H-----Q-E-----P-V--Y----- 446  
DB 3839 RKNGHLLNVQDNQNSYRLLSSGSLVLIISPSVDDTATYECTVTNGAGDDKRTVLTVOVPP 3898  
QY 447 SVA-----F-S-----P--D----- 453

Db 3899 SIADPTDFLWTKHAPAVITCTAGVFPFPIHWTGIRLLPRGDGVRILSSGAIELAT 3958  
Qy 454 ---GRY---L---ASG--S- 461  
Db 3959 QLNHAGRYTCVARNAGSAHRHVLHVHPPVQPPQPSSEHVLNNPILLPCEATGPSP 4018  
Qy 462 F---D---KCV--- 466  
Db 4019 FITWQEGINWNTSGRNHVLPSGGLQISRAVEDAGTYMCVAQNPAGTAGIKLKNVQV 4078  
Qy 467 ---H---I-W---NT-Q--- 472  
Db 4079 PPVISPHLKEYVIAVDKPIITLSCADGLPPPDITWHKDGRAIVESIRQVRLSSGSLQIAF 4138  
Qy 473 V---C---L---HY--- 477  
Db 4139 VQPDAGHYTCMAANVAGSSSTSTKLTVHVPPIRSTEGHYTNENSQAILPCVADGPT 4198  
Qy 478 ---L-N---GQ---VL---L---N--- 485  
Db 4199 PAIWKKNVLLANLLGKYTAEPYGLILENVLEDSGFYTCVANNAAAGETHVSLTVH 4258  
Qy 486 ---L-G---R-SIC---L-YT---LP-H---H-IV--- 500  
Db 4259 VLPTFTLPGDVSINKGEQLRLS-CRATGIPPLKLTWTFNNIIPAHFDSVNGHSELVIE 4317  
Qy 501 ---V---V---I-PL--- 505  
Db 4318 RVKEDSGTVCTAENSVGFPVKATGFVYVKEPPVFKGDYPSNWIPLGNGNAILNCEVKGD 4377  
Qy 506 ---AL---IE---L---L-VL---K 514  
Db 4378 PPTIQNRKGVDEISHRIQLNGSLAIGTVNE 4413

## RESULT 8

US-10-408-765A-1895  
; Sequence 1895, Application US/10408765A  
; Publication No. US20040101874A1  
; GENERAL INFORMATION:  
; APPLICANT: Ghosh, Soumitra S.  
; APPLICANT: Fahy, Eoin D.  
; APPLICANT: Zhang, Bing  
; APPLICANT: Gibson, Bradford W.  
; APPLICANT: Taylor, Steven W.  
; APPLICANT: Glenn, Gary M.  
; APPLICANT: Warnock, Dale E.  
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION  
; FILE REFERENCE: 660088.465  
; CURRENT APPLICATION NUMBER: US/10/408, 765A  
; CURRENT FILING DATE: 2003-04-04  
; NUMBER OF SEQ ID NOS: 3077  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1895  
; LENGTH: 5636  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: VARIANT  
; LOCATION: 3003, 3041, 3367  
; OTHER INFORMATION: Xaa = Any Amino Acid  
US-10-408-765A-1895

Query Match 71.2%; Score 2596.5; DB 16; Length 5636;  
Best Local Similarity 9.7%; Pred. No. 2.7e-29;  
Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;  
Qy 1 M-S---I---SS---D---EV--- 8  
Db 1 MISWEVHTVFLPALLYSLAQDASPQSEIRAEFFPEGASTLAFVFDVTGSMYDLDLVQVI 60

Qy 9 ---NF-IV--- 12  
Db 61 EGASKILETSLKRPKPLFNFALVPPHDPPIGPTTTTDPKFOYELRELYVQGGDCPE 120  
Qy 13 ---Y---R---Y---LQ--- 17  
Db 121 MSIGAIAKIALEISLPGSFYVFTDARS KDYRLTHEVQLIQKOSQVVFVLTGDCDRTH 180  
Qy 18 ---E---SG-F---S-H-S---A---F 26  
Db 181 IGKVVETASTSSGQVHLDKKQVNEVLKWVEEAVQASKVHLLSTDHLSQAVNTWRIPP 240  
Qy 27 ---T---FG---I---K 31  
Db 241 DPSLKEVTVSLSGSPMIEIRNPLGLIKKFGHLLNIHNSAKVNVVKEPEAGMWTVK 300  
Qy 32 ---S-H---I---S-Q--- 36  
Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDKKTYSRPVQGIPTYVLLNTSGISTPARID 360  
Qy 37 ---S---NI---N---G--- 41  
Db 361 LLELLSISGSLKTI PVKYPHRKYGIWNISDFVPNEAFFLKVGTGYDKDDYLFQVSS 420  
Qy 42 ---A---L---VP---P---A 47  
Db 421 VSFSSIVDPAPKVTMEKTPGYVLPQGPICSDVSLPFTLSFVRNGVTGLGVDQYLKESA 480  
Qy 48 ---A---L---I---S-IIQ---KG--- 56  
Db 481 SVSLDIKVTLSDEGYECIAVSAGTGRAQTFDVSPPVTVQVNNVTVTGCEAVLT 540  
Qy 57 ---L-Q---Y---V-E--- 61  
Db 541 CLIIISAVNLTWQRNDRVRLAEPARIRTLANLSLELKVENDAGEYHCVSSEGGSS 600  
Qy 62 -A---EVS1---NE--- 68  
Db 601 AASVFLTVQEPKVTVMKNSQFTGSGSEVSIKSGATGPKPKIAWTVNDMFIVGSHRYRM 660  
Qy 69 ---DGLTF---D---G---R---P---I--- 78  
Db 661 TSDGTLFIKNAAPKADAGIYCLASNSAGTDKQNSTLYRIEAPKLMVQVSELLVALGDITV 720  
Qy 79 -E---S---L-S-LI---DA---V--- 87  
Db 721 MECKTSGIPPPQVWPKGDLRLRPFTLIIDPLLLGLLKIQETQDLADGYTCTVAINESGR 780  
Qy 88 ---M---P-DV---VQ---T--- 94  
Db 781 ATGKITLDVGSPPVFIQEPADYSMEIGSNVTLPCYVQGYPEPTIKWRLLDNMDFISRPFS 840  
Qy 95 ---R--- 95  
Db 841 VSSISQRLTGALFILNLWASDKGTICEAENQPKIQSETTVTGTGLVAPLIGISPSVAN 900  
Qy 96 ---QQ---A---Y---R---D-KL--- 103  
Db 901 VIEGQQLTLPCTLLAGNPERRWIKNSAMLLQNPYITVRSQGLSHERVOLQDGGEYTC 960  
Qy 104 ---A---Q-QQ---AA--- 109  
Db 961 VASNVAGTNNKTSVVVHVLPFIHQOQILSTIEGIPVTLPCKASGNPKPSVWSKKGEL 1020  
Qy 110 ---AAA---A-A---A-AA---A--- 118  
Db 1021 ISTSSAKFSAGAGSLVYVSPGESGEVYCTATNTAGYAKRKVQLTVYVPRVFGDLRG 1080  
Qy 119 -SQ---Q---GS--- 123  
Db 1081 LSQDKPVEISVLAGEBVTLPCEVKSUPPIITWAKETQLISPFSPRHTFLPSGSMKITET 1140  
Qy 124 ---A---K---N--- 126

Db 1141 RTSDSGMYLCVATNIAGNVTAQVAKLVNVHPPKIQRGKHLKVQVQORVDIPCNQAQGTPLP 1200  
QY 127 -----GE-----N-----T--A-N-G----- 133  
Db 1201 VITWSKGGSTMLVDGHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEILHVQE 1260  
QY 134 -----E-----E-N-N-----GA-----H----- 139  
Db 1261 PPTVELEPPYNTTFORVANQRIEPPCPAKGTPKPTIKWLHNGRELGTGREPGISILEEG 1320  
QY 140 T-----IA-----N-----N-----HT-----DM----- 148  
Db 1321 TLLVIASTVPDNGEYICVAVNEAGTERKYNLKVHVPVPIKDKEQVSNVSLNLQTLN 1380  
QY 149 -----M-----E----- 150  
Db 1381 FCEVEGTPSPIIMWYKDNVQVTSSTIQTVNNKILKLFATPEDAGRYSCAKAINAGTS 1440  
QY 151 -----VD-----G-----DV--E-----I-----P-----PN 160  
Db 1441 QKYNIDLVPPPTIIGTNPPEKVSUVRNDRDALECOVKGTPTPPDIHWFKDGKFLGDPN 1500  
QY 161 -----K-A-----V----- 165  
Db 1501 VELLDRGOVLHLKNARENDKGRYQCTVSNAAGQAQDKLTIYIPPSIKGNVTTDISVL 1560  
QY 166 -----RG-----H----- 168  
Db 1561 INSILKLECKTGLPMPAITWYKDGQPIIMSSQALYIDKQYLHIIPRAQVSDSATYTCV 1620  
QY 169 -----E-S-----E-----VFI-----C-A-----W----- 177  
Db 1621 ANVAGTAESFHVDDVVPVPIEGNLATPLNKQVIAHSLTLECNAGNPSPIITWLDGV 1680  
QY 178 -----N-----P-V--SD-- 182  
Db 1681 PVKANDFRIEAGGKLEIMSAQEIIDRGQYICVATSVAGEKEIKEYVDVLVPPAIEGGDE 1740  
QY 183 -----LL-----ASGS-----G-----D-----S-- 191  
Db 1741 TSYFVIVNNLLELDCHVTGSPPTIMWLKDGQLIDERDGFKILLNGRKLVIQAQVSN 1800  
QY 192 -----TA-----R-----I-W 196  
Db 1801 GLYRCHMAANTAGDHKEFEVTVHVPPTIKSSGLSERVVVYKPVALQCIANGIPNSITW 1860  
QY 197 -----NL-----S 199  
Db 1861 LKDDQPNTAQGNLKIQSGRVLQIAKTLLDAGRYTCVATNAAGETQOHIQLHVHEPPS 1920  
QY 200 -E-----N-----S-----T-----SGST----- 207  
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLSGTSMTFLNRQIIDIES 1980  
QY 208 -Q-----L--V-----LR-H-----C-----I-- 215  
Db 1981 AQISDAGIYKCVAINAGATELFYSLOVHVAPISGSNNMVAVVNNPVPLECEARGIPA 2040  
QY 216 -----R-----E-----G-----G--Q--D-- 221  
Db 2041 PSLTWLKDGSVPVSNGLQVLSGRILALTSTQISDTGRYTCVAVNAAGEKQRDIDLVR 2100  
QY 222 -VPSN-----KD-----VT-----S-L- 231  
Db 2101 YVPPNINGEQNVSVLISQAVELLQSDAIPPTPLTWLKDGHPLLKPKGLSISENSVLK 2160  
QY 232 -----D-----W--N--S-----EG--TLL-- 240  
Db 2161 IEDAQVODTGRYTCEATNVAGTEKKNNVNIWPPNIGSGDELQTLTVIEGNLISLCE 2220  
QY 241 AT-----GS--Y-D-----G-----F-----A- 249  
;

Db 2221 SSGI PPPNLWKKGSPVLTD SMGRVRLSGRQLQISIAEKSDAALYSCVASNVAGTAK 2280  
QY 250 -----R-----I-----WKDG----- 256  
Db 2281 KEYNLQYIRPITNSGSHPTBIIIVTRGKSISECEVQIIPPTVTWMDGHPHAKGV 2340  
QY 257 -----NL-AS-----T-----L-----GQH----- 265  
Db 2341 EILDEGHILQAKNIHVSOTGRYCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPENISVV 2400  
QY 266 -K-----G-PI-----FALK--W----- 274  
Db 2401 EKNSVLTCEASGIPLPSTWF--KDGWPVLSNSVRILSGGRMLRMOTMTWEDAQYTC 2458  
QY 275 -----N-----K-K-----GN----- 279  
Db 2459 VVRNAGEBERKIFGLSVLPVPHVGTLEDVKVKEKQSVTLTCEVTGNPVPBITWHKDG 2518  
QY 280 -----F--I-----L-----S--AGV-- 286  
Db 2519 QPLQDEAHHIIISGGRFLQITNVQVPHGTGRYTCLASSPAGHKSRSPSLNVFVSPTIAGVG 2578  
QY 287 -D--K--T-----T--I-----I-W----- 293  
Db 2579 SDGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLSNRNRILPGRTLQILNAQ 2638  
QY 294 -D-----A-----H-----TG-----E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVYIPIINKGLWPGLSPEKVKIKVNTLTLECE 2698  
QY 300 A-----K-----Q-----Q----- 303  
Db 2699 AVAIPASLSWYKDGQPLKSDDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758  
QY 304 F-----P-F-----H- 307  
Db 2759 FDNVIOVPPSFQKWEIGNMLDTRNGEAKDVIIINNPISLYCEETNAAPPTLTWYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315  
Db 2819 LTSSDKVILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878  
QY 316 -----WO-----SN----- 319  
Db 2879 LPBEVTVLNKSALIECLSSGSPAPRNSWKDQOPLLEDHFKFLNGRILQILNTQITD 2938  
QY 320 -----NT-----FAS--C-----S 326  
Db 2939 IGRYVCVAENTAGSAKKYFNLNVHVPVPSVIGPKSENLTVVNNFISLTCEVSGFPDPLS 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKXQPIKLTNTLIVPGRTLQIIRAKVSDGGEYTCIAINXAGESKKKFSLTVVYVPPS 3058  
QY 332 --H-----V--C-----K-----L--GO-- 338  
Db 3059 IKDHDSELSVNVNREGTSVLECESNAVPPVITWYKNGRMITESTHVEILADGQMLHI 3118  
QY 339 -----D-----R-----P-IK----- 343  
Db 3119 KKAESVDTGOYVCRAINVAGDDKNFNLNVYVPPSIEGPEREIVETISNPVTLTCDATG 3178  
QY 344 --T-----F-----Q--G-----H-----T----- 349  
Db 3179 IPPTTIAWLKNYKRIENSLSLEVRILSGSKLQIARSQSDSGNYTCIASNMEGKAKYY 3238  
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
Db 3239 FLISIQVPPSVAGAEIPSDVSVLGENVELVCNANGIPTPLIOWLKDGPKTASGETERIV 3298  
QY 361 --G--NLL-A--S-----C--S-----D-----D-----MTL-- 373  
Db 3299 SANGSTINIYALTSDTKYTCVATNPAGEEDRIFNLNVYVPTIRGNKDEAKLMTYVD 3358

[illegible]

RESULT 9  
US-10-464-368-69  
; Sequence 69, Application US/10464368  
; Publication No. US20040023356A1  
; GENERAL INFORMATION:  
; APPLICANT: Krumlauf, Robb  
; APPLICANT: Ellies, Debra  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION  
; FILE REFERENCE: 40716-IP-017  
; CURRENT APPLICATION NUMBER: US/10/464,368  
; CURRENT FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/388,970  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 140  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 69  
; LENGTH: 4599  
; TYPE: PRT  
; ORGANISM: MOUSE  
US-10-464-368-69

Query Match 71.0%; Score 2589.5; DB 15; Length 4599;  
Best Local Similarity 9.3%; Pred. No. 1.3e-29;  
Matches 428; Conservative 65; Mismatches 14; Indels 4095; Gaps 370

QY	1	MS	-----I-----S-----	4
Db	1	MSQELLAITLGLLPNAEVLIVGANOQHLCDPGEFLCHDHVTCVSQSWLGDGDPDCPD	60	
QY	5	SD-----EV-----N-----F-----	10	
Db	61	QSDBSLTCPEBEVKEICPLNHIACHGSSACVHLSKLCNGWDCPDGDFEGHGCHELLPSC	120	
QY	11	--L-----V-----Y-----R-----Y-L-Q-----	17	
Db	121	QQLNCQFCAMVRNATRCYCEGFEVAEDGRCKDQDECSYIGCSQTCNTYGSVACSC	180	
QY	18	ESGF-----S-----H-----SAP-T--F--Gik-----S-----H-----	33	
Db	181	VE-GVIMQSDNRSCVKXGHEPTDKAPMLLSILETIEFLVINGSKMTLSSARNETHLTD	239	
QY	34	-----I-S-QS-----	39	
Db	240	FIYSEMICWIESRESSNLKCGQITKAGRLTDORIIINSLOSQFNVEQMAFDWLTRNIYF	299	
QY	40	-----NG-----A--L-----V-----	45	
Db	300	VDHVSDRIFVCNFGNSVCVTLIESELHNPKAIAADPIAGKLFPTDYGNVPKIERCDLGM	359	
QY	46	-----P-A-AL-----I-----SIQKGLQ-----Y--59		
Db	360	NRTRIVYSKAEQPSALALDLVNRLVYWDLYDVGVVDYQSKNRHTIVQ-CRQVKHLXG	418	
QY	60	--V-E-----AEVS-----IN-----E-68		
Db	419	ITVPEDYLYA-TSSDNFNIIRINFGTDIHSIIKMESARGIRTYOKRTQPTVRSHACEV	477	
QY	69	D-----G-----TL-----F-----DGR-----76		
Db	478	DAYGMPGCSCHILLSSSYKTRTCRCRTGFNMGSDGRSKPKNELFLFYGKRGPIVRG	537	
QY	77	-----PIE-----SL-----S-LI-----84		
Db	538	MDLNTKIADCEMIPIENLVNPRALDFHAEANYIYPADTTSPFLGRQIDGTERILKDD	597	
QY	85	D-----AV-----M-----P-----89		
Db	598	LDNVEGIAVDWIGNNLYTNDGHRKTINVARLEKASQSRKTLLEGMSHPRAIVDPVNG	657	
QY	90	-----D-V-----V-Q-----93		
Db	658	WNYWTDWKEDKIDDSVGRIEKAMGDGVNRQVFTVSKMLWPNGLTLDFTSTLYWCDAAYD	717	

QY 94 -----T-R-----Q----- 96  
Db 718 HIEKVLNGTHRKVVYSGKELNHPFGLSHGNVFWFTDYNNGSIFQDLMTNEVTLRHE 777  
QY 97 -----QAY--R--D-K-----LA----- 104  
Db 778 RAPLGLQIYDPRKQGDNNCRINNGCGGTLCLAIAGRVACADNQLLDENGTTCTFNP 837  
QY 105 -----Q-----Q-----Q----- 107  
Db 838 EEIRFHCKEPEFRCKNKHCIQARWKCDGDDCLDGSDEDSVTCFNHSCPDQDFKCQNNR 897  
QY 108 -----A-----AA--AA-----AA----- 116  
Db 898 CIPKRWLDCDAGNCGNEDESNOTCTARTCOAQDFSCGNGRCIPTAWLDCDREDDCGDQTD 957  
QY 117 -AAS-----Q-----Q-----Q----- 121  
Db 958 EVASCEPPTCEPLTQFICKSGRCISNKHWCDDTDDCCDRSDEVCVHSCLDQFRCSSGR 1017  
QY 122 ---GS-A---KN-G---E---N-T---A-----N-----GE-134  
Db 1018 CIPGHWACDGDNDGDFSDETHINCTKEARSAGCIGNEFQCRPDGNCIPDLWRCDGEK 1077  
QY 135 -----E---NGA---H-----T---I----- 141  
Db 1078 DCBDGSEKGCNGTIRLCHDKTKFRCRSTGRGINNAWVGDDVDCBQDEBDCDSFLCG 1137  
QY 142 -----AN-----NH----- 145  
Db 1138 PPYPCANDFSVCLOPEKLCNGKDCPDGSDGDLCDCESLNNGCSNHCSSVVGRIVC 1197  
QY 146 -----T---D-----MM---E-----VD-152  
Db 1198 SCPEGHOLKDNRTCEIVDYCASHLRCQVCEQKHMKVCKSCYEGWALGTGDSCTSVDS 1257  
QY 153 -----GD----- 154  
Db 1258 FEAPLIFSIRHETRIDLHKGDYSLVPLGRLNTIALDFHFNQSLLYWTDVVEDRIYRGKL 1317  
QY 155 -----VE-----I----- 157  
Db 1318 SESGVSIAIEVVEHGLATEGLTVDMWAGNIYWIDSNLDQIEVSKLDGSLRATLAGAM 1377  
QY 158 --P-----P-----N-----K----- 161  
Db 1378 EHPRATLDPYGLIFLWTDWDANFPRIEASMSGACKTIYKDMKTGAMPNGLTVDFHER 1437  
QY 162 -----AV-----VLRGHE-----SEV----- 172  
Db 1438 RIVWTDARSADIAYSFVDTGNMIEIRHGHEYLSHPPFAVSLYSEVYVWTDWRTWTLAKAN 1497  
QY 173 -----F-----I-----CA--176  
Db 1498 WTCQNVSVIOKTSAQFPDLQIYHPSRQAPNPCAANEGRGPCSHLCLINHNESAACAP 1557  
QY 177 -----W-----NP-----VS---D--- 182  
Db 1558 HLMKLSDDKTKCYEMKKFLLYARRSEIRGVIDPNVYFITAFTVPDIDVAVIDEASE 1617  
QY 183 --LL-----A---SG-----S-G--D-----S---T--- 192  
Db 1618 ERLYWTDIKTQITRAPINGTGLETVISRDIQISIRGLAVDWVSRNLYWISSEFDEQINV 1677  
QY 193 AR-----I-----W-----N----- 197  
Db 1678 ARLDGSLKTSIIHGDKPQCLAAHPVGKLYWTDGNTINMANWGDGNSKILFQNKQEPVG 1737  
QY 198 LS-----S---T-----SG-----S-----T----- 207  
Db 1738 LSIYVYENKLYWISSGNGTINRCNLGDNLEVIEMKEELTKATALTIMDKLWADQNL 1797

QY 208 -OL-----VLR-----H----- 213  
Db 1798 AQLGTCKNKEDGRNPSILRNKTSQVVMKYYDKAEQOQSNQCVNNGGCQCLPTSETTR 1857  
QY 214 -----C-----IR----- 216  
Db 1858 TCMCTVGYVLQKNRMSQGIIESFLMYSVHEGIRGIPLEPRDKVDALMPISGAFAVGIDF 1917  
QY 217 -----EG----- 218  
Db 1918 HAENDTIYWDMLGNKISRAKRDQWTKEDVVTNGLRVEGIAVDWITAGNIYWTDHGFNLI 1977  
QY 219 -----GO-----D-----V 222  
Db 1978 EVARLNGSFRYVLIISQGLDQPRSIAPHPEKFLFWTEWQVPCIGIKARLDGSEKWMIVSV 2037  
QY 223 ---P---S---NK-----D---V---T-----S-LD----- 232  
Db 2038 GITWPNGISIDYBENKLYWCARDARKIERIDLTGANREVLVSGSNVDLFSVAVFGAIY 2097  
QY 233 W-----N-S-----E-CT----- 238  
Db 2098 WSDRAHANGSVRRGHKNDATEVTWRTGLGVNLKEIKIFNRVREKGTNVCAKENGCCQL 2157  
QY 239 -----LLA---T-----G-----S---Y--245  
Db 2158 CLYGRNSRRTACAHGYLAGDGVTCRLRHEGYLLYSRTILKSIHLSDETNLNSPVRPYEN 2217  
QY 246 -----D---G---F---A-----R---I----- 251  
Db 2218 PNVFKNIIALAFDYNORREGTNRIFYSDAHFGNIQLIKONWEDROQVIVENVGSGEGLAVH 2277  
QY 252 -----W-----TK--D---G-----NL----- 258  
Db 2278 RAWDTLYWTSSTSTSTRHTVDTQTRFPAIDREAVITMSBDDHPHVALDECQNLFWTNW 2337  
QY 259 -----A---ST-----LG-----Q-264  
Db 2338 NEQHPISMRATLTGKNAHVVSVDILTPLNGLTIDHRAEKLYFSDGSLGKIERCEYDSQR 2397  
QY 265 H--K-GP-----IF-ALKW-----NK-----K----- 277  
Db 2398 HVIKSGPGTFLSLAVDYSYIFWS-DWGRRAILRSNKYTGGETKILRSIPHQMGIIV 2456  
QY 278 -----G-----N-----F 280  
Db 2457 ANDTNSCELSPCALLGGCHDLCLLPDGRVNCSCRGDRVLLANNRCVTKNSSCNIFYSEF 2516  
QY 281 -----I-----L-----S-AG-----V----- 286  
Db 2517 ECGNGDCVDYVLTCDGI PHCKDKSDEKLYLCENRSCSGFKPCYNRRCPVPHGLCDGTND 2576  
QY 287 -----D-K---T-----T-I----- 291  
Db 2577 CGDSSDELCKVSTCSTVEFRCADGTICPRSARCNQNMDCSDASDEKGNNTDCTHYKL 2636  
QY 292 -----I-----W-----D-----AH----- 296  
Db 2637 GVKSTGFIKRNSTSLCVLPSWICDGNDCGDYDELKCPVQNKHKEENYFGCPSGRCIL 2696  
QY 297 -T-----G-E-----A-----K----- 301  
Db 2697 NTWVCGQKDCEDGLDELHCDSSCSWNQFACSVKVKCISKHWICDGEDDCGDSLDESISIC 2756  
QY 302 -----Q-----Q-----F----- 304  
Db 2757 GAVTCAADMFSQCGSHACVPQHWLDCGERDCPDGSDLSAGCAPNNTCDENAFMCHNKV 2816  
QY 305 --P--F--H-----S-----AP-----ALD-----VD--WQ-----S 318  
Db 2817 CIPKQVCDHDDDCGSGSDEFLOCGYRQCGPBEFRCA-DGRCLVNTLMQCDGDFDCPDSS 2875  
QY 319 ---N-----NT--FASC-----STD-MC-I-----HV--C--- 334





[illegible]

Db	1138	PPKYP	CANDTSV	CLQPEK	LNGRKDC	PCDGS	DEGDL	CDECS	LNNGG	CNSH	CVVPG	RGI	VC	1119
Qy	146	---	T	---	D	---	MM	---	E	---	VD	---	152	
Db	1198	SCPEGH	QKKNFT	CEIVDY	CASHLR	CSQVCE	QOKEHV	KSCVCE	GWALGT	DGESC	TS	VD	1257	
Qy	153	---	GD	---	---	---	---	---	---	---	---	---	154	
Db	1258	FEAFI	PSIRHE	IRIDL	HKGDY	SLVPL	GLRNT	IALD	PHFNQ	SLLY	WTVD	VEDR	1317	
Qy	155	---	VE	---	---	---	---	---	---	---	---	---	157	
Db	1318	SESGV	SAIEVV	VEHGL	ATPEG	LTV	WIA	NIW	IDS	NLD	QIEV	KDGS	1377	
Qy	158	---	P	---	N	---	---	---	K	---	---	---	161	
Db	1378	EHPRA	LDPY	GILF	WTWD	ANFP	RIES	ASMS	GAGK	TIY	KDM	KTG	1437	
Qy	162	---	AV	---	VLR	GHP	---	---	SEV	---	---	---	172	
Db	1438	RIVMT	DARSD	AIYAF	YDGT	NMIEI	IRGHE	YLSH	PF	AVS	LYG	SEV	1497	
Qy	173	---	P	---	---	---	---	---	---	---	---	---	176	
Db	1498	WTGON	SVIO	KTS	QA	PPDL	QIYH	PSR	QOP	ANP	CA	NEGR	1557	
Qy	177	---	W	---	NP	---	---	---	VS	---	D	---	182	
Db	1558	HLMKL	SDKKT	CEYK	MKFL	YARR	SHIR	GV	DI	NPY	NF	IT	1617	
Qy	183	---	LL	---	A	---	SG	---	S	---	D	---	192	
Db	1618	ERLYW	TDKT	OTIT	RA	FN	GT	LE	T	IS	R	D	1677	
Qy	193	AR	---	I	---	W	---	---	N	---	---	---	197	
Db	1678	ARLDG	SLK	SI	HH	DKP	QCLAA	HP	VR	GK	LY	WT	1737	
Qy	198	LS	---	EN	---	S	---	T	SG	---	S	---	207	
Db	1738	LS	DY	VEN	KLY	WISS	NGT	IN	R	C	LN	D	1797	
Qy	208	QL	---	---	VLR	---	H	---	---	---	---	---	213	
Db	1798	AQLG	T	CN	KR	GR	NPS	IL	R	NK	TSG	V	1857	
Qy	214	---	---	C	---	IR	---	---	---	---	---	---	216	
Db	1858	TCM	CTGY	LQ	KNR	MSC	Q	G	I	ES	FL	MS	1917	
Qy	217	---	---	---	EG	---	---	---	---	---	---	---	218	
Db	1918	HAEND	TI	WT	D	M	L	N	K	I	S	R	1977	
Qy	219	---	---	---	Q	---	---	---	---	---	---	---	222	
Db	1978	EVAR	L	NGS	FR	Y	I	S	Q	L	D	P	2037	
Qy	223	---	P	---	S	---	NK	---	D	---	V	---	232	
Db	2038	GITW	P	NG	I	S	I	D	E	N	K	L	2097	
Qy	233	W	---	N	---	S	---	---	E	---	GT	---	238	
Db	2098	WSDRA	H	ANG	S	VR	R	G	H	K	N	D	2157	
Qy	239	---	---	LLA	---	T	---	G	---	---	S	---	245	
Db	2158	CLYR	G	NS	R	RT	C	A	C	A	G	Y	2217	
Qy	246	---	---	D	---	G	---	F	---	A	---	R	251	
Db	2218	P	N	F	K	N	I	A	L	F	D	N	2277	

QY 252 -----W-----TK--D--G-----NL----- 258  
D 2278 RAWDTLYWTSSSTSSITRHTVDQTRFGAIDREAVITMSBDDHVLALDEBQNLMTW 2337  
QY 259 -----A-----ST-----LG-----Q- 264  
D 2338 NEQHPIMRATLTGKNAHVWVSTIITPNGLTIDHRAEKLKLYFDGSLGKIERCEYDGSOR 2397  
QY 265 H--K-GP-----IF-ALKW-----NK-----K----- 277  
D 2398 HVIKSGPGTFLSLAVDSYIFWS--DWGRAILRSNKYTGGETKIILRSIDIPHOPMGIIV 2456  
QY 278 -----G-----N-----P 280  
D 2457 ANDTNSCELSPCALLGGCHDLCLLTPDGRVNCRCGRDVRLLANNRCVTKNSSCNIYSEF 2516  
QY 281 -----I-----L-----S-AG-----V----- 286  
D 2517 ECGNGCDVYVLTCDGIPHCKDKSDEKLLYCENRSCRGFKPCYNRRRCVPHGKLCDDTND 2576  
QY 287 -----D-K--T-----T-I-----K----- 291  
D 2577 CGDSSDELCKVSTCTVFRCADGTICIPRSARCNQNMDCSDASDEKGCNNTDCTHFKYL 2636  
QY 292 -----I-----W-----D-----AH----- 296  
D 2637 GVKSTGFIKRNSTSLCVLPFSWICDGSNDGDSYDELKCPQNKHKCEENYFCGPGRCIL 2696  
QY 297 -T-----G-E-----A-----K----- 301  
D 2697 NTWVCDGQKDCEDGLDELHCDSSCNQFACSVKCKISKHWICDGBDDCGDSLDESISIC 2756  
QY 302 -----Q-----Q-----F----- 304  
D 2757 GAVTCADMFQCGSHACVPHQWLCDGERDCPDGSDLSAGCAPNNTCDENAFMCHNKV 2816  
QY 305 -P--F--H-----S-----AP-----ALD-----VD--WQ-----S 318  
D 2817 CIPKQVCDHDDCGDGSDEFLQCGYRQCPEFRCA--DGRCLVTLWQCDGDFDCPDSS 2875  
QY 319 -----N-----NT--FASC-----STD-MC-I-----HV--C----- 334  
D 2876 DEAPINPRCSAEHSCNSFFM--CKNGRCIPS--DGLCDIRDGCGSDETNCHINECLSK 2933  
QY 335 KL-G-QD-R--PI-----KT-----F-----Q-----G-----H- 348  
D 2934 KISGCSQDCQDLPVSYKCKWPGFQKDKGKTCVDIDECSSGFPSCQOCINTYGYKCHC 2993  
QY 349 -----T-----N----- 350  
D 2994 ABGYETQDPNPGCRSLSDPEPLIADQHEIRKISTDGSNTYLLKQGLNNVIALDPDYR 3053  
QY 351 -E-----V-NA-----I-K--W-D----- 358  
D 3054 EBFYIWDSSRNGSRINRMCLNGSDIKVYHNTAVPNALAVDWIGKLYWSDTEKRIIEV 3113  
QY 359 -----PT-----GNLL-ASC----- 367  
D 3114 SKLNGLYPTVLVSKRLKFRDLSDPRAGNLWIDCCYEPHIGRVGMGTNOSVVIETKI 3173  
QY 368 S-----D-----M-----TL-K-I-W----- 376  
D 3174 SRPMALTIDYVNHRLYWADENHIEFNSMDGSHRHKVPNQDIPGVIALTLFEDYIYWDGK 3233  
QY 377 -----S-MK-----Q-D-----N-----C----- 383  
D 3234 TKLSRVHKTSGADRLSLNSWHAITDIQVYHSYRQPDVSKELCTVWNGGCSHLCLLPG 3293  
QY 384 -VH-----D-----L----- 387  
D 3294 KTHTCACPTNFYLAADNRTCLSNCTASQPRCKTDKCIPIFWWKDVTDDCGDGSDEPDPC 3353

QY 388 -----Q-----Q-----H-----NKE-I-- 394  
D 3354 EFKCQPGRFQCGTGLCALPAFICDGENDCGNSDELNCDTVCLAGQFKCTKNKKIPVN 3413  
QY 395 -----Y-----T-----I-K-W-----SP----- 401  
D 3414 LRCNGQDDCGDEBEDEKOCPENSCSPDYFOCKTKKICISKLWVWDEDPDCADASDEANCDK 3473  
QY 402 -T-GP-----G-----T-N--N--P----- 409  
D 3474 KTCGPHFOCKNNCIPDHWRCNQNDQNSDSDNCKPOTCTLKDPLCSNGDCVSRFW 3533  
QY 410 -----N-----A----- 412  
D 3534 CDGEFPCADGSDBKNCETSCSKDQFQCSNGQCLSAKWKCDGHEDCKYGEKNCPEAPPV 3593  
QY 413 -----LMLA-----SAS-----F----- 420  
D 3594 CSSEYMCASGGCLSAKLCNKEPCDVGSDDEMCVIECKEPOFQCKNKAYCIPIRWLCD 3653  
QY 421 -----D-S--T-----V-R-----L--W-----D--V-- 429  
D 3654 GIYDCVDSDEETCGGSGICRDBEPLNNSLCKLHFW/CDGEDDCGNSDEAPDMCVKF 3713  
QY 430 -----DR-----GI----- 433  
D 3714 LCPPTPRYCRNDRICLQLEKICNGINDCGNSDEHSCGKLSKPKCKBFTCSNRN 3773  
QY 434 CI-----HT-----L-----T----- 439  
D 3774 CIPMELQCDLDDCGDGSDEOGLKTPIEHTCENNNGPCDDAYCNOIKTSVFCRCKPGF 3833  
QY 440 -----K-----H-----OE----- 443  
D 3834 QRNMKGRECADNECLLFGICSHHCLNTRGSKVCVCDQNFQKNNSCIAKSGEDQALYIA 3893  
QY 444 -----P-----V-----Y----- 446  
D 3894 NDTDILGVYFVYPNYSGHQIISHVEHNSRI/TGMDVHYQRNVITWSTQFNPGGIFYKMWIDA 3953  
QY 447 -----S-----VA-----FS----- 451  
D 3954 REKQANSGLICPEFRPRDIAVDWVAGNVYWTDSHSMHFSYTYTHWTSRLSYINVGOL 4013  
QY 452 -----P-----DG----- 454  
D 4014 NGPNCRTLNTNAGEPYAIVNPKGMVYTVIGDHSHEEBAAMDGTLLRRVLVQKNLQRP 4073  
QY 455 -----R-Y--LA--S--GS----- 461  
D 4074 TGLTVDFHGERIYWADFELSIIIGSVLYDGSPPVSVSSKQGLLHPHRI/DVFEDYIYGAP 4133  
QY 462 -----F-----DK----- 464  
D 4134 KNGIFRVQFGHGSVEVLALGVDKTSILVSHRYKQLNLPCLDLSCDPLCLNPSGAT 4193  
QY 465 CV-----H-----I-----W-----N-- 470  
D 4194 CICPEGYMMGTCHDDSLDSDCKLTCENGRCILNEKGLRCHCWPSYSGRCVNHHC 4253  
QY 471 T--Q-----VC-L-----H-----Y----- 477  
D 4254 SNYCQGGTCIPSTLGRPTCICALGFTGPNCKGKAVCEDSCHNGGSCVVTAGNOPYCHCQA 4313  
QY 478 -LNG--Q--V--L--L--N-----L-G-----R----- 488  
D 4314 DYTGRQCYVYCHHYCVNSESCTIGNDGSVECVCPTRYEGPKCEIDKVCRRGHGHCIIINK 4373  
QY 489 -----S--I--C-L--Y----- 493  
D 4374 DNEDIFCNCTNKGIASSCQDCGYCYNGTGCOLDPETSIPVCVCTNWSGTQGERPAPKS 4433  
QY 494 -----TL-----P 496

Db 4434 SKSEHISTRSIAIIVLVLVLTTLVTLVGLVVCRRKRTKTRRQPIINGINVEIGNP 4493  
QY 497 -----H-H-----L-----V-----I-----PL-----504  
Db 4494 SYNMYEVDHSDGGLLEFBSFMDPVKSRVYIGSSGSAFKLPHTAPPIYLSNLDKGLPTFG 4553  
QY 505 -----V-A-----L-I-----ELL-----V-LK 514  
Db 4554 PTNYSNPVYAKLYMDQNCNCRNSLASVDERKELLPKKIEIGIR 4595

## RESULT 11

US-10-120-801-74  
; Sequence 74, Application US/10120801  
; Publication No. US20030203843A1  
; GENERAL INFORMATION:  
; APPLICANT: Pena, Carol  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Padigar, Muralidhara  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spyttek, Kimberly  
; APPLICANT: Mehraban, Fuad  
; APPLICANT: Topper, James N.  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Wasserman, Scott  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Smithson, Glennada  
; APPLICANT: Gunther, Erik  
; APPLICANT: Komuves, Laszlo  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-340  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR FILING DATE: 2001-05-03  
; PRIOR FILING DATE: 2001-05-16  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 74  
; LENGTH: 5175  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-120-801-74

Query Match 70.9%; Score 2588; DB 14; Length 5175;  
Best Local Similarity 8.8%; Pred. No. 2.5e-29;  
Matches 453; Conservative 49; Mismatches 8; Indels 4660; Gaps 389;

QY 1 M-----SI-----SS-----D-----E-----7  
Db 1 MGRSPSWLYGVGLLLATTSSVNDKNDPTGKSSLAFFVDITGSMFDDLVQVREGAAK 60  
QY 8 -----V-----N-----F-----L-----VY-----13  
Db 61 IFKTVMAQREKLIYINIMVPHDFYLGELINTTDTYFMRLSKVYVHGGDCPEKTLTG 120  
QY 14 -----R-----Y-----L-----QE-----SG-----F-----SHSAF-T 27  
Db 121 ILKALQISLPSSFTYVFTDARSKDYHLEDEVLNTIQEKSSVWFVMTGDCGNRTHPGFRT 180

QY 28 -----FG-I-----KS-----H-ISO-----S-NI-----39  
Db 191 YEKIAAASFGVFLHLEKSDVSTVLEYVRHAVKQKVHLMYEAREGGTVSRNIPVDKHL 240  
QY 40 -----N-----GA-----LV-----P-P-----46  
Db 241 ELTISLSDKSDNDLIVLRDPGRTVDKRLYSKEGGTIDLKNVLRIRLKDPSPGVWTV 300  
QY 47 -----A-----A-----LI-----S-I-----52  
Db 301 NTNSLKHTIRVFGHGAVDFKYGASRPLDRIELARPPVLNQDTYLLINNTGLIPPTV 360  
QY 53 -----I-----Q-----KGL-----Q-Y-----VE-----AE-----63  
Db 361 GEIDLVDYHGLSLYKAVASPHRTNPNMYFAGFPVPPKGLFFRVQGYDNEVEFMRIAPT 420  
QY 64 -----V-----S-I-----N-----ED-----GTLF-----73  
Db 421 AIGSVIVGGPRAFMSPHQEFVGRDLNLCTVESASAYTIYWKVTGEDIIGPLFYHNTD 480  
QY 74 -----D-----G-----R-----P-I-----E-----79  
Db 481 TSVWTIPELSLKDAGEYECRVISNNGNYSVKTRETRESPPEIFGVRNVSVPLGEAAFLH 540  
QY 80 -----S-----L-----SL-I-----DA-----86  
Db 541 CSTRSAGEVEIRWTRYGATVFNPNPTNPTNGTLKIHVTRADAGVYECMARNAGCMST 600  
QY 87 -----VM-----P-DV-----V-----92  
Db 601 RKMRLDIMEPPSVKVTQDVYFNMREGVNLSCAMGDPKPEVHWYFKGRHLNDYKQVG 660  
QY 93 -----T-----R-----Q-----QAYRD-----KLA-----Q-----105  
Db 661 QDSKFLYIRDTATHDEGTGYECRAMSQQA-RDITDLMLATPPKVEIIONKMMVGRGDRV 719  
QY 106 -----Q-----Q-----A-----A-----A-----A-----111  
Db 720 SPECKTIRGKPHKIRWFKNGKDLKPDYIKINEGQLHMGAKDEBAGAYSCVGNMAG 779  
QY 112 -----A-----A-----A-----A-----A-----A-----114  
Db 780 KDQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEMQKGNVLLATLNPP 839  
QY 115 -----A-----A-----A-----ASQ-----120  
Db 840 RYTQADGNLLITDAQIEDQGOFTCIARTYQQSQSSTLMVTGLVSPVLGHVPPEQLI 899  
QY 121 -----Q-----GS-----123  
Db 900 EGQDLTLCVVVLGTGPKPSIWMKDDKPVKEGPTIKIEGGGSLRLRGNPKDEGKYTCI 959  
QY 124 -----A-----K-----N-----N-----GE-----128  
Db 960 AVSPAGNSTLHINVLIIKKPEFYKPEGGIVFKPTISGMDEKHAHVAVNSTHVDLDEGFA 1019  
QY 129 -----NT-----A-N-----G-----133  
Db 1020 IPCVSGTTPPIITWLDGRPITPNSRDTFTVADNTLIVRKADKSYSGVYTQATNSAGD 1079  
QY 134 -----E-----E-----N-----136  
Db 1080 NEQKTTIRIMTNPISPGQSFNMVVDLFTIPCDVYGDGPKFVITWLLDDKPFTEGVNNE 1139  
QY 137 -----G-----AH-----TI-A-N-----N-----144  
Db 1140 DGSLLTIPNVNEAHRGFTTCHQNAAGNDRTVTTLVHTTPTINAEQEKIALQNDIVLE 1199  
QY 145 -----HT-----D-----MME-----V-----D-----152  
Db 1200 CPAKALPPVRLWYEGEKIDSLIPIHTIREDGALVQNVKLENTGVFCQVSNLAGEDS 1259  
QY 153 -----G-DV-----EIP-----P-----NK-AV-----163

Db 1260 LSVTLTVHEKPKIIESEVGVVVDVVGFTTIBIPCRATGVPVIRTNKNGIDLMDKEKFS 1319  
QY 164 V-----LR-----G-----H-----E-----S----- 170  
Db 1320 VDNLGLTIRIYEADKNDIGNVNCVVTNEAGTSQMTTHVDVQEPPIILPSTQNTTAVUGDR 1379  
QY 171 -----E-----V-----F-----I-----C----- 175  
Db 1380 VELKCYVEASPPASVTFRRGIAIGTDTKGYYVESDGLTVIQSASVEDATIYTCASNPA 1439  
QY 176 -----A----- 176  
Db 1440 GKAEANLQVTIASPIKODPVVTQBSIKESHPSLYCPVFSNPLPQISWYLNKPLIDD 1499  
QY 177 -----W----- 178  
Db 1500 KTSWKTSDDKRLHVFKAKITDSGVVKCVARNAAEGSKSFQVEVIVPLNLDSEYKKKV 1559  
QY 179 -----PVS----- 181  
Db 1560 FAKEGEEVTLGCPVSGFPVQPINNVVDGTWVEPGKVKYKATLSNDGLTLHFDSDSVSKQEG 1619  
QY 182 -----D-----L-----LA----- 186  
Db 1620 NYHCVAQSGNILDIDVELSVLAVPIVGEDDNLEVLGKDISLSCDLQTESDDKTTFFWS 1679  
QY 187 -----GS-----G-----DS-----T-----A-----R----- 194  
Db 1680 INGESDRPDNVQIPSDGHRLYTIDAKPENNGKCMCRVTNSAGKABRTITLQVLEPPV 1739  
QY 195 -----EN----- 199  
Db 1740 EPVFEANQKLGNNPIILQCVTGPKPTVWIKIDGNDVDKSWLPDESLSLLRIEKLTKG 1799  
QY 200 -----EN----- 199  
Db 1800 SAQISCTAENKAGTASRDFFIQIAAFTFKNEGDOETIFRESEITITLDCPVLGDFQITW 1859  
QY 208 -----Q-----L-----VL-----R-----H-----C----- 216  
Db 1860 MKQGLPLTENDAIFLTLNTRLTILNANRDHEDIYTCVANTAGQVSKDFDVVVQVLPKIK 1919  
QY 217 -----EG-----G-----Q-----DVP----- 223  
Db 1920 NAVVTLIINEGEIEILTCDAGNPTTAKWDFNQGLPKAEVFNHNHTVWVNNVTKYHT 1979  
QY 224 -----SNK-----DV-----T-----S-----L-----D----- 232  
Db 1980 GYVKYATNKVQAVKTINVHRTKPRFESGLTESELTVNLTRITLCECDVDDAIGVGIS 2039  
QY 233 W-N-----SE-----G-----TL-----L-----A-----T-----GSY----- 245  
Db 2040 WTVNGKPLAETDGVQTLAGRFLHIVSAKTDHDSYACTVTNEAGVATKTNLNFVQVPP 2099  
QY 246 -----DG----- 248  
Db 2100 TIVNEGEYTVIENNSVLVPCVETGKPNVPTWKDGRPVGDLKSVQVLSQEQFKIYHA 2159  
QY 249 -----A-----R----- 250  
Db 2160 EIAHKSYSICMAKNDVGTABISFDVDIITRPMIQIKNIIVTAIKGALPFKCPIDDDKN 2219  
QY 251 -----IW----- 257  
Db 2220 FKQIILWRNYQPIDLEADARITLSNDRRLTILNVTENDEQVSCRKNDAGENSFDF 2279  
QY 258 -----L-----A-----S-----TL----- 265  
Db 2280 KATVLPPTIIMLDKDKNTAVEHSVTLSCTATGKPEPDIITWFKDGEAIHIENTADIIP 2339  
QY 266 ----- 265

Db 2340 NGBELNGNQLKITRIKEGDAGKYTCADNSAGSVQEDVNVNVNITIPKIEKDGPDSYESQQ 2399  
QY 266 -----K-----G-----P----- 269  
Db 2400 NERVVISCVPYARPPAKITWLKAGKPLQSDKFVKTSANGOKLYLFKLRETDSSKYTCIAT 2459  
QY 270 -----F-----A-----L-----K-----W----- 274  
Db 2460 NEAGTDKDPKVSMLVAPSFDEPNIVRRITVNSGNPSTLHCPAKGSPSPITITWLKDGNAI 2519  
QY 275 -----N-----K----- 276  
Db 2520 EPNDRYVFFDAGRQLQISKTEGSDQGRYTCIATNSVSGDLENTLVBIIPPVIGDERREA 2579  
QY 277 -----K-----G-----N-----FI-----LS----- 283  
Db 2580 VAVIEGSELSFCDNSNSTGVDEWQXGLTINQDTLGRGDSFQIPSSGKMSFSLARKSD 2639  
QY 284 -----AG-----V-----DK-----T-----TI-----I----- 292  
Db 2640 SGRYTCIVRNPAEARKLFDFAVNDPPSISDELSSANIQTIVPYYPVEINCVVSGSPHPK 2699  
QY 293 -----W-----D-----A-----H-----T-----G----- 298  
Db 2700 VYWLFDKPLEPDSAAAYELTNNGETLKIVRQVEHAGTYTCEAQNNGKARKDLVRVTA 2759  
QY 299 -----E-----AK-----Q----- 302  
Db 2760 PPHFEKEREVARVGDTHMLTCNAESSVPLSSVYVWHAHDESQNGVITSKYAANEKTLN 2819  
QY 303 -----Q-----F----- 304  
Db 2820 VTNIQLDDSGFYCTAVNEAGITKFKFLIVETPYFLDQOKLYPIILGKRLTDCSATG 2879  
QY 305 -----P-----F----- 306  
Db 2880 TPTPTILFMKQKRLNESDEVDIIGSTLVIDNPQKEVEGYTCIAENKAGRSEKMMVEV 2939  
QY 307 -----H----- 308  
Db 2940 LLLPKLSKEWINVEVQAGDPLTLECPEDTSGVHTWSRQFGKQGLDMRAQSSSKSL 2999  
QY 309 -----A-----P-----A-----L-----D-----V-----D----- 315  
Db 3000 YIMQATPEDADSYSCIAVNDAGAEAVFQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRT 3059  
QY 316 -----W----- 317  
Db 3060 EGIPPEISWFLDGKPILEMPGVYTKQGLSLRIDNIKNQBGRYTCVAENKAGRAEQDT 3119  
QY 318 -----S-----N-----N-----N-----T-----F----- 322  
Db 3120 YVEISEPPRVNASEVMRVVEGRQTTIRCEVFGNPEPVVNWLDKGPYTSDDLQFSTKLS 3179  
QY 323 -----A-----SC-----S-----TD-----N-----C----- 330  
Db 3180 YLHLRETTLDGTYTCIATNKAGESQTTDVEVLVPPRIEDEERVLQKQEGTYNVHQC 3239  
QY 331 -----I-----HV-----C-----K----- 335  
Db 3240 VTGRPVYVTKRNGKEIEQFPNVLHIRNATRADEGKYSIASNEAGTAVADFLIDVFTK 3299  
QY 336 -----L-----Q-----D-----RP-----I-----K----- 343  
Db 3300 PTFETHETFTNIVEGESAKIECKIDGHPKPTISWLKGGPFNNMDNIIILSPRGDTMLKA 3359  
QY 344 -----T-----F-----Q----- 346  
Db 3360 QRFDGGLYTCVATNSVGDSEQDFKNNVYTKPYVIDETIDOTPKAVAGGEIILKCPVLGNPT 3419  
QY 347 -----G-----HT-----N-----E-----VN-----A-----I----- 355  
Db 3420 PTVTWKRGDDAVPNDNRHTIVNNYDLKINSVTTEDAGQYSCIADVNEAGNLTHYAAEVIG 3479

QY 356 -K- - - - -W- - - - -D- - - - -P- - - - -T- 360  
Db 3480 KPTFRKGNLYEVIENDTTIMDCGVTSRPLSPISWFRGDKPVYLDRYSISPDGSHITI 3539  
QY 361 -G- - - - -N- - - - -L- - - - -LA- - - - -S- - - - -C- - - - -S 368  
Db 3540 NKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKDKSNIIGNPLAIVARTIYLECPIS 3599  
QY 369 -D- - - - -D- - - - -M- 371  
Db 3600 GIQPDVITWTKMGMDINMTDSRVILAQNNETFGLNVQVTDQGRYTCATNRGKASHDF 3659  
QY 372 -TL- - - - -K- - - - -I- - - - -W- - - - -S- - - - -M- - - - - 378  
Db 3660 SLDVLSPEFDIHGTQTIKREGDTITLTCPIKLAEDIAQVMDVSVTKDSRALDGLTD 3719  
QY 379 -K- - - - -Q- 381  
Db 3720 NVDISDDGRKLTISOASLENAGLYCTIALNRAGEASLEFKVEILSPPVVIDISRNDVQPV 3779  
QY 382 -N- - - - -C- - - - -V- - - - -H- - - - -D- - - - -L- - - - -Q- - - - -H- - - - - 390  
Db 3780 AVNQPTIMRCVATGHPSPSIKMLKNGKEVTDDEINIRIVEQGVQLILRTDSDHAGKWSV 3839  
QY 391 -N- - - - -KE- - - - -I- - - - -Y- - - - -TI- 399  
Db 3840 AENDAGVKELEMLVDVFTPPVSVKSDNPICALGETITLFCNAGNPPYQPKWAKGSLI 3899  
QY 400 -SP- - - - -T- - - - -G- - - - -P- - - - -GT- 407  
Db 3900 FDSFDGARISLKGARLDIPLHKKTVDGDTYCOALNRAAGTSEASVVDVLVPPINRDGID 3959  
QY 408 - - - - -N- 409  
Db 3960 MSPRLPAQSLTLQCLAQGPVQMRWTLNGTALTHTPGITVASDSTFTQINNVSLSDK 4019  
QY 410 - - - - -N- - - - -A- - - - -NLM- 417  
Db 4020 GVTYCAENVAGSDNLMYVVDVQAPVISINGTKQVIEGELAVIECLVEGYPAPQVSWLR 4079  
QY 418 - - - - -A- - - - -SPD- 421  
Db 4080 NGRVETGVQVRYVTDGRMLTIEARSLDSGIYLSATNEAGSAQOAYTLEVLVSPKII 4139  
QY 422 -ST- - - - -VR- - - - -L- - - - -W- - - - -DV- 429  
Db 4140 TSTPGVLTSPSSGKFSPLCAVGVDPDPIISWTLNGNDIKDGENGTIGADGTLHIEKABE 4199  
QY 430 - - - - -D- 431  
Db 4200 RHLIYEKAKNAGADTLEFPVQTVAPKISTSGNRYINGSEGTETVIKIESESSEFS 4259  
QY 432 -G- - - - -I- 434  
Db 4260 WSKNGVPLLSNNLIFSEDIYKILSTRLSQGEYSCTAANKAGNATQKTNLVGAPK 4319  
QY 435 I- - - - -H- - - - -TL- 440  
Db 4320 IMERPRTQVVKGDQVTLWCEASGVQPAITWYKDNELLTNTGTDEATTATTKKSVIFSSI 4379  
QY 441 - 440  
Db 4380 SPSQAGVYTCKAENWVASTEEDIDLIVMIPPEVVPVPMNVSTNPTQTVFLSCNATGIPEP 4439  
QY 441 - 442  
Db 4440 VISWRDSNIAIQNEKYQILGTLTALRNVLPPDDGFYHICIAKSDAGCKIATKRLVKNKP 4499  
QY 443 - 446  
Db 4500 SDRPAPIWVECKGPKKTEYIMIDRGDTIPDNPQLLPWKDVEDSSLSNGSIAYRCMPGPR 4559

QY 447 -S- - - - -V- - - - -A- - - - -F- - - - -S- - - - -P- - - - -DG- - - - - 454  
Db 4560 SRTVLLHAAPQFIYVVKPKNTTAAIGAIVELRCSAAGPHTTITWAKDGKLIEDSKFEIAY 4619  
QY 455 - 454  
Db 4620 SHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAITCYER 4679  
QY 455 - - - - -R- - - - -Y- - - - -LA- - - - -S- - - - -G- - - - - - - - - - - - - - - - 460  
Db 4680 NOAYSRLTWXNGVPMKPLAGIHFPMNNGSLVILDTSSLKEGDLLEYTKVNRHRSI 4739  
QY 461 - - - - -S- - - - -P- - - - -DK- - - - -C- - - - -V- - - - -H- - - - -IW- - - - - 469  
Db 4740 PHLTSAFEGVPEVKTIDKVEVNGDSVILDCVTSPLTHVVVTKNDQKMLDDDAIYVL 4799  
QY 470 - 469  
Db 4800 PNNSLVLLNVEKYDEGVYKCVASISGKAFDDTQLNVYGGSSRRREAYKKENEDASTTIT 4859  
QY 470 - 472  
Db 4860 TTSPTTTTTELTETITIIIPALITLPAKOYPTDDYHEGSANDDGGFTTQDSLFEFNPPLH 4919  
QY 473 - - - - -V- 476  
Db 4920 PEISVNTTCAGTINENGDCVOKDKTHNLKILTGENHCPEGFAMPHTRICEDLDECAF 4979  
QY 477 Y- - - - -LN- - - - -G- - - - -QVLLN- - - - -LG- - - - -R- - - - -SI- - - - -C- - - - -L- - - - - 492  
Db 4980 YQPCDFECINYDGGFQ- - - - -CNCPLGYELABEGCRDVNECESVRCECDGKACFNQLGGYECID 5037  
QY 493 - - - - -YTL- 500  
Db 5038 DPCPANYSLVDDRYCEPECENCTSTPIQVHMLAIPSLPISHIATLTAYDKSGRVLNDDT 5097  
QY 501 -VI- - - - -PL- 510  
Db 5098 YALSDGAFGLARGMTSGPFTIKAVKRGHAQVWTRVLAAGDHHKVRVRAHSDHATNELH 5157  
QY 511 - - - - -LVL 513  
Db 5158 APKETNPLVL 5167

## RESULT 12

US-10-369-493-6859  
; Sequence 6859, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6859  
; LENGTH: 5175  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-6859

Query Match 70.9%; Score 2588; DB 14; Length 5175;  
Best Local Similarity 8.8%; Pred. No. 2.5e-29;  
Matches 453; Conservative 49; Mismatches 8; Indels 4660; Gaps 389;







441 QY ----- 440  
4380 Db SPSOAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMVSTNPNROTIVFLSCNATGIPEP 4439  
441 QY ----- 442  
4440 Db VISWRDSNTAIONNEKYQILGTTTLAIRNVLPDDGGFYHCIAKSDAGQKIATKLIWKP 4499  
443 QY ----- 446  
4500 Db SDRPAPIWECDEKGPKEKTEYIMIDRGDTPDDNPQLLPWKDVEDSSLSNGSIARCMGPGR 4559  
447 QY -S-V-A-F-----P-----DG----- 454  
4560 Db SSRTVLLHAPOFIVKPKNTTAAIGAIVELRCSAAGPPHTITWAXDGKLIEDSKFEIAY 4619  
455 QY ----- 454  
4620 Db SHLKVTLNSTSDGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAITCYER 4679  
455 QY -R-Y-LA-----S-G----- 460  
4680 Db NOAYSRLTWEYNGVMPKPKLAGIHFMNNGSLVILDTSSLSKEGDELYTCKVNRRRHSI 4739  
461 QY -S-F-----DK-----C-V-----H-TW----- 469  
4740 Db PHLTSAFEGVPEVKTIDKVEVNGDSVLDCEVTSPLTTHVVTWKNDQKMLDDDAIYVL 4799  
470 QY ----- 469  
4800 Db PNNSLVLLNVEKYKCVASNSIGKAFDDTQLNVYGGSSRRAYKKENEDASTTTIT 4859  
470 QY ----- 472  
4860 Db TTSPTTTTETPLTTIIPALITLPAKQYPTDDYHEGSANDDGGFTTQSLPEFNPPLH 4919  
473 QY -V-----C-----L-----H----- 476  
4920 Db PEISVNTDCAGTINENGCDVKDKGTHNLKILTGENHCPEGFAMNPHTRICEDLDECAF 4979  
477 QY -LN-G-QVLLN-LG-----R-SI-C-----L----- 492  
4980 Db YQPCDFECINYDGGFQ-CNCPGLGYELABEGCRDVNECESVRCEKACFNOLGGYECID 5037  
493 QY -YTL-----P-H-----H-L-----V----- 500  
5038 Db DPCPANYSLVDDRYCEPECENTSTPIQVHMLAIPSGLPISHIATLITAYDKSGRVLNDTT 5097  
501 QY -VI-----PL-----V-A-----L-----I-----EL- 510  
5098 Db YALSDTGAPLARGMTSGPPTIKAVKRGHAQVWTVNRVLAAGDHHKVRVRAHSDHATNELH 5157  
511 QY -----LVL 513  
5158 Db APKRTNLFVL 5167

## RESULT 13

US-10-369-493-6861  
; Sequence 6861, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6861  
; LENGTH: 5175  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-6861

Query Match 70.9%; Score 2588; DB 14; Length 5175;  
Best Local Similarity 8.8%; Pred. No. 2.5e-29;  
Matches 453; Conservative 49; Mismatches 8; Indels 4660; Gaps 389;

QY 1 M-----SI-----SS-----D-----E----- 7  
Db 1 MGRSPSWLYGVGLGALLLATTCSVNDKNDPTKSSLAFAFVFDITGSMFDDLQVREGAAK 60  
QY 8 -V-----N-----F-L-VY----- 13  
Db 61 IFKTMAQREKLIYNYIMVPHDPYLGEINTTSTYFMRQLSKVYVHGGDCPEKTLTG 120  
QY 14 -R-Y-----L-OE--SG--F-----SHSAF-T 27  
Db 121 ILKALQISLPSSFIYVFTDARSKDYHLEDEVLTIOEKSSVVFVMTGDCGNRTHPGFRT 180  
QY 28 -FG-I---KS-----H-ISO-----S-NI----- 39  
Db 181 YEKIAAASFGQVHLEKSDVSTVLEVVRHAVKQKVHLMAYEARERGGTVSRNIPVDKHL 240  
QY 40 -N-----GA-----LV-----P-P----- 46  
Db 241 ELTISLGDKDDNDLIVLRDPEGRTVDKRLYSKEGGTIDLKNVKLIRLKDPSGVWTV 300  
QY 47 -----A-----A-----LI-S-I----- 52  
Db 301 NNSRLKHTIRVFGHGAVDFKGFASRDLRIELARPRVLPVNDOTVLLINMTGLIPPGTV 360  
QY 53 -I-----Q-----KGL-----Q-Y--VE---AE- 63  
Db 361 GEIDLVDYHSHLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYEDNIEFMRAPT 420  
QY 64 -V-----S-I-----N-----ED---GTLF----- 73  
Db 421 AIGSVTVGGPRAFMSPIHQEFVGRDLNCLSTVESASAYTIYWKGTGEDIIGPLFVHNTD 480  
QY 74 -----D-----G-----R-P-I-----E----- 79  
Db 481 TSVWTIPELSLKDAGEYECRVISNNGNYSVKTRETRESPEIFGVRNVSVPLGEAAFLH 540  
QY 80 -S-----L-----SL-I-----DA----- 86  
Db 541 CSTRSAGEVEIRWTRYGATVFNPNPTNPTNGTLKIHVTRADAGVYECMARNAGMST 600  
QY 87 -VM-----P-DV-----KLA-----Q-----V- 92  
Db 601 RKMRLDIMEPPSVKVTQDQVYFNMRBGNLSCEAMGDPKPEVHWYFKRHLNDYKYQVG 660  
QY 93 Q-----T--R--Q--QAYD---KLA-----Q----- 105  
Db 661 QDSKFLYIRDAATHDEGTYECRAMSQQA-RDITDMLATPPKVEIIONKMMVGRDRV 719  
QY 106 -----Q-----Q-----A-A-A-A-----A- 111  
Db 720 SPECKTIRKPHKIRWFKNGKDLKPDYIYKINEQLHMGAKDEADAGYSCVGENMAG 779  
QY 112 -A-----A-----A-----A-----A----- 114  
Db 780 KDQVANLSVGRVPTTIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNNP 839  
QY 115 -A-----A-----A-----ASQ----- 120  
Db 840 RYTQLADGNLLITDAQIEDQGOFTCIARNTYQQSQSSTLMTVGLVSPVLGHVPPBEQLI 899  
QY 121 -Q-----GS----- 123

Db 900 EGQDLTSCVVLGTPKPSIWIKDDKPVBEPTIKIEGGSLRLRGNPKDEGKYTCI 959  
QY 124 ---A---K---N---GE---128  
Db 960 AVSPAGNSTLHINVLIKKPEFVYKPEGGIVFKPTISGMDEKHAVVNSTHVDLGEFA 1019  
QY 129 ---NT---A-N-G-133  
Db 1020 IPCVSGTPPIIITWYLDGRPIITPNSRDFVTADNTLIVRKADKSYGVVTCOATNSAGD 1079  
QY 134 -E- ---E-N-136  
Db 1080 NEQKTTIRIMNTMISPGOSSFMVVDLFTIPCDVGDPKPVIWLLDDKPTFEGVNE 1139  
QY 137 -G- ---TI-A-N-144  
Db 1140 DGSLLTIPNVNEAHRGFTTCHAOAAGNDTRVTLTVHHTPTINAENOEKIALQNDIVLE 1199  
QY 145 ---HT---D---MME---V---D-152  
Db 1200 CPAKALPPVRLTYEKEKIDSQIIPHITREDGALVQNKLENTGVFCQVSNLAGEDS 1259  
QY 153 ---G-DV---EIP---P---NK-AV-163  
Db 1260 LSYTLTVHEKPIISEVPGVVDVVGFTIIPCRATGVPEVIRTNKNGIDLQWDEKPS 1319  
QY 164 V---LR---G---H---E---S---170  
Db 1320 VDLGLTRIYEADKNDIGNYCVVTVNEAGTSQMTTHVDVQEPPIILPSTQNTNNAVVGDR 1379  
QY 171 ---E---V-F---I-C-175  
Db 1380 VELKCVYASPPASVTWFRGIAIGTDTKGYVVEDTLVIQASVEDATYITCKASNPA 1439  
QY 176 ---A---176  
Db 1440 GKAEANLOVTVIASPDKDPDVTVQBSIKESHPSLYCPVFSNPLQIISWYLNKPLIDD 1499  
QY 177 ---W---N---178  
Db 1500 KTSWKTSDDKRLHVFKAKITDSGVYKCVARNAAGEKSKSFQVEVIVPLNLDSEKYYKKV 1559  
QY 179 ---PVS---181  
Db 1560 FAKGEBVLGCPVSGFPVQINWVDGTVVEPKKYKGATLNDGLTLHFDVSVKQEG 1619  
QY 182 ---D-L-LA---S-186  
Db 1620 NYHCAQSKGNILDDIVELSLAVPIVGEDDNLVFLGKDISLSCOLQTESDDKTTFVMS 1679  
QY 187 ---GS---DS---T---A-R-194  
Db 1680 INGSERPDNVQIIPSDGHLXYITDAKPNNGKYMCRVTNSAGKABRTTLTDLVLEPPVFV 1739  
QY 195 ---EG---IW---N---LS-199  
Db 1740 EPVFEANQKLIIGNPIILOQVGTGNPKPTVIWKIDGNDVDKSWLFDSELSLLRIEKLTKG 1799  
QY 200 ---EN---S---T---S-GS---T-207  
Db 1800 SAQISCTAENKAGTARDFFIQNIAPTFKNEGQETIIPRESEITILDCPVLGDFQITW 1859  
QY 208 ---Q---L-VL---R-H---C---IR-216  
Db 1860 MKQGLPLTENDAITLNTLRLTILNANRHEDIYTCVANNTAGQVSKDFVWVQVLPKIK 1919  
QY 217 ---EG---G---Q-DVP-223  
Db 1920 NAVVLEINEGEBEILTCDAEGNPTTAKWDFNQGLPKAEAVFNNNHTVVVANNVTKYHT 1979  
QY 224 ---SNK---DV---T-S-L-232

Db 1980 GVVKYATNKVGOAVKNTINHVTRKPRPESGLTESELTVNLTRSTILECDVDDAIGVGIS 2039  
QY 233 W-N---SE-G-TL---L---A-T---GSY-245  
Db 2040 WTVNGKPFIAETDGTQTLGAGRFLHIVSAKTDHSHSYACTVTNEAGVATKTNLNFVQVPP 2099  
QY 246 ---DG---F-248  
Db 2100 TIVNEGGEVTVIENNSLVLPCEVTGKPNPVVTVTKDGRPVGLKSVQVISEGQOFLVHA 2159  
QY 249 ---A---R-250  
Db 2160 EIAHKGSIYMAKNDVGTAEISFDVDIITRPMIQKIGKNIKVIATKGAALPFPKPIDDDKN 2219  
QY 251 ---IW---T---K-D-G-N-257  
Db 2220 FKQIILWLNYPIDLEADARITRLSNDRRLTILNVNTEDEGQYSCRVKNDAGENSFDF 2279  
QY 258 ---L---A-S-TL---GO-H-265  
Db 2280 KATVLPPTIIMLDKDKNTAVEHSTVILSCPATGKPEPDITWFKDGEAIHIENIADIIP 2339  
QY 266 ---K-G-P-269  
Db 2340 NGELNGNQLKITRIKEDGAKYTCEADNSAGSVEQDVNVNVTIPKIEKDGPIDYESOQ 2399  
QY 266 ---K-G-P-269  
Db 2400 NERVISCPVYARPPAKITWLKAGLPQSKDFKVTKSANGOKLYFLKLRDTSKYTCIAT 2459  
QY 270 ---F---A---L---K---W-274  
Db 2460 NEAGTKDRFKVSMVAPSPDEPNVRRITVNSGNPSTLHCPAKGSPSTIWLKDGNAI 2519  
QY 275 ---N---K-276  
Db 2520 EPNDRYVFFDAGRLQISKTEGSDQGYTCIATNSVSDSDLENTLEVIIPPVIGDERREA 2579  
QY 277 ---K-G-N-283  
Db 2580 VAVIEGSELFCDSNSTGVVWQKGLTINQDTLURGSFQIIPSSGKMSFLSARKSD 2639  
QY 284 ---AG---V---DK-T-TI---I-292  
Db 2640 SGRYTCIVNPAGEARKLFDFAVNDPPSISDELSSANIQTIVPYVPEINCVVSGSPHPK 2699  
QY 293 ---W---D-A---H---T---G-298  
Db 2700 VYWLFDKPLEPDSAAAYELTNNGETLKI VRSQVEHAGTYTCEAQNNGKARKDFLVRVTA 2759  
QY 299 ---E---AK---Q-302  
Db 2760 PHFEKERBEVAVRGDTMLLTCAESSVPLSVYHMHDESQVNGVITSKYAANEKTLN 2819  
QY 303 ---Q---F-304  
Db 2820 VTNIQLDDEGFFYCTAVNEAGITKFFKLI VIETPYFLDQKLYPIILGKRLTDCSATG 2879  
QY 305 ---P---F-306  
Db 2880 TPEPTILFMKGKRLNESDEVDIGSTLVIDNPQKEVEGRYTCIAENKAGRSKDMMEV 2939  
QY 307 ---H---S-308  
Db 2940 LLPPLKSKWINVEVQAGDPLTECPEDTSGVHITWSRFQKQGLDMRAQSSDKSL 2999  
QY 309 ---A-P-A-L---D---V-D-315  
Db 3000 YIMQATPEDADSYSCIAVDAGGAFAVQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRT 3059  
QY 316 ---W-317  
Db 3060 EGIPPEISWFLDGKPILEMPGVYTKQGLSLRIDNKPQEGRYTCVAENKAGRAEQDT 3119

QY 318 -----S-----N-----N-----T-----F----- 322  
Db 3120 YVEISEPPRVWASEVMRVEGRQTTIRCEVFGNPEPVNWLKDGPEYTSDDLQFSTKLS 3179  
QY 323 -----A-----SC-----S-----TD-----M-----C-----330  
Db 3180 YLHLRETTLADGGTYTCIATNKAGESQTTTDEVLVPPRIEDEERVLOQKEGNTVMVHCQ 3239  
QY 331 -----I-----HV-----C-----K-----335  
Db 3240 VTGRPVVYVWRNGKEIEQFNPVLRNATRADEGKYSCIASNEAGTAVADFLIDVFTK 3299  
QY 336 -----L-----GO-----D-----RP-----I-----K-----343  
Db 3300 PTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRFPFNWDNIIILSPRGDTLMILKA 3359  
QY 344 -----T-----F-----Q-----346  
Db 3360 QRFDDGLYTCVATNSYGDSEQDFKVNVTYKPYIDETIDQTPRAVAGEIILKCPVLGNPT 3419  
QY 347 -----G-----HT-----N-----E-----VN-----A-----I-----355  
Db 3420 PVTWKRGDPAVNSRHTIIVNNDLKINSVTEDAGQYSCIAVNEAGNLTHYAAEVIG 3479  
QY 356 -----K-----W-----D-----P-----T-----360  
Db 3480 KPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYDRYSISPDGSHITI 3539  
QY 361 -----G-----N-----L-----LA-----S-----C-----S-----368  
Db 3540 NKAKLSDGKGYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPTS 3599  
QY 369 -----D-----M-----371  
Db 3600 GIQPDVIVTKNGMDINMTDSRVLQONNETGNIENVQVTDQRYTCTATNRGKASHDF 3659  
QY 372 -----TL-----K-----I-----W-----S-----M-----378  
Db 3660 SLDVLSPPFEDIHGTQPTIKREGDTITLTCPIKLAEDIAQVMDVSWTKDSRALDGLTD 3719  
QY 379 -----K-----Q-----D-----381  
Db 3720 NVDISDGRKLATISQASLENAGLYTCIALNRAGEASLEFKVELSPVIDISRNDVQPOV 3779  
QY 382 -----N-----C-----V-----H-----D-----L-----Q-----Q-----H-----390  
Db 3780 AVNQPTIMRCVATGHPPFSIKWLKNGKEVTDDENIRIVEBQQVQLQILRTDSDHAGKWSV 3839  
QY 391 -----N-----KB-----I-----Y-----TI-----KW-----399  
Db 3840 AENDAGVKELEWLVDFTPPVSVKSDNPICKALGETITILFCNAGNPYPQLKWKAGGSLI 3899  
QY 400 -----SP-----T-----G-----P-----GT-----N-----407  
Db 3900 FDSPDGARISLAGARLDIHLKKTVDGYTCQALNAAGTSEASVVDVLVPPPEINRDGID 3959  
QY 408 -----N-----P-----409  
Db 3960 MSPRLPAQOSLTQCLAQKPVQMRWTLNGTALTHTSTPGITVASDSTFIQNNVSLSDK 4019  
QY 410 -----N-----A-----NLM-----LA-----S-----417  
Db 4020 GUYTCYAENVAGSDNLMYVVDVQAPVINSNGTKQVIEGELAVIECLVEGYPAQVSWLR 4079  
QY 418 -----A-----SPD-----421  
Db 4080 NGNRVETGVGVYVTDGRMLTIIERSLDSGIYLCSATNEAGSAQAYTLEVLVSPKII 4139  
QY 422 -----ST-----VR-----L-----W-----DV-----429  
Db 4140 TSTPGVLTPSSGSKFSLPCAVRGYPDPPIISWTLNGNDIKDGENGHTIGADGTLHIEKABE 4199

RESULT 14

US-10-120-801-75

; Sequence 75, Application US/10120801

; Publication No. US20030203843A1

QY 430 -----D-----R-----431  
Db 4200 RHLIVECTAKNDAGADTLEFPVQTIIVAPKISTGNRYINGSEGTETVIKCEIESSEFS 4259  
QY 432 -----G-----I-----C-----434  
Db 4260 WSKNGVPLPSNNLIFSEDYKLIKILSTRLSDQGEYSCTAANKAGNATOKTNLNVGVAPK 4319  
QY 435 -----H-----TL-----TK-----440  
Db 4320 IMERPRTQVHHKGDQVTLWCEASGVPOPAITWYKONELLTNTGVDETATTKKSVIFSSI 4379  
QY 441 -----441  
Db 4380 SPSQAGVYTCKAENMVASTEEDIDLIVMIPPEVPERMVSTNPRQTVFLSCNATGIPEP 4439  
QY 441 -----H-----Q-----442  
Db 4440 VISMRDSNIAIQONNEKYQILGTTLAINVLPPDDGFYHCAKSDAGOKIATRKLIIVKP 4499  
QY 443 -----E-----P-----V-----Y-----446  
Db 4500 SDRPAPIWECDEKPKKTEYMWIDRGDTPDNDPQLLPWKDVEDSSLSNGSIAYRCMPGPR 4559  
QY 447 -----S-----V-----A-----F-----S-----P-----DG-----454  
Db 4560 SSRTVLLHAAPOFIVKPKNTTAAIGAIVELRCSAAGPPHTITWAKDGKLIEDSKPEIAY 4619  
QY 455 -----R-----Y-----LA-----S-----G-----460  
Db 4680 NOAYSRGLTWEYNGVMPKPKLAGIHFMMNGSLVILDTSSLKGGDLELYTCKVNRRRHSI 4739  
QY 461 -----S-----F-----DK-----C-----V-----H-----IW-----469  
Db 4740 PHLTSAFEGVPEVKTIKDVENVNGDSVVLDCVTSPLTTHVVTWTKNDQKMLDDDAIYVL 4799  
QY 470 -----470  
Db 4800 PNNSLVLLNVEKYDEGVYKCVASNSIGKAFDDTQNLNVYGGSSRRREAYKKENEDASTTIT 4859  
QY 470 -----N-----TO-----472  
Db 4860 TTSPTTTTETPLTTIIPALITLPAKQYPTDDYHEGSANDDGFPGTQDSLFEFNPPLH 4919  
QY 473 -----V-----C-----L-----H-----476  
Db 4920 PEISVVNTDCAGTINENGDCVDKDGKTHNLKILTGENHCPEGFAMNPHTRICEDLDECAF 4979  
QY 477 Y-----LN-----G-----QVLLN-----LG-----R-----SI-----C-----L-----492  
Db 4980 YQCPDFECINYDGGFQ--CNCPLGYELABEGCRDVNECSVRCECDGKACFNQLGGYECID 5037  
QY 493 -----YTL-----P-----H-----H-----L-----V-----500  
Db 5038 DPCPANYSLVDDRYCEPECENCTSTPIQVHMLAIPSGLPISHIATLTAYDKSGRVLNDTT 5097  
QY 501 -----VI-----PL-----V-----A-----L-----I-----EL-----510  
Db 5098 YAISTDGAPLARGMTSGPFTIKAVKRGHAQVWTRNVLAAAGDHHKVRVRAHSDHATNELH 5157  
QY 511 -----LVL-----513  
Db 5158 APKETNFLVL 5167

GENERAL INFORMATION:  
APPLICANT: Pena, Carol  
APPLICANT: Guo, Xiaojia  
APPLICANT: Shimkets, Richard  
APPLICANT: Padigar, Muralidhara  
APPLICANT: Kekuda, Ramesh  
APPLICANT: Spytek, Kimberly  
APPLICANT: Mehraban, Fuad  
APPLICANT: Topper, James N.  
APPLICANT: Malyankar, Uriel  
APPLICANT: Wasserman, Scott  
APPLICANT: Edinger, Shlomit  
APPLICANT: Smithson, Glennda  
APPLICANT: Gunther, Erik  
APPLICANT: Komuves, Laszlo  
TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 21402-340  
CURRENT APPLICATION NUMBER: US/10/120,801  
CURRENT FILING DATE: 2002-04-11  
PRIOR APPLICATION NUMBER: 60/285748  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: 60/286068  
PRIOR FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 60/286292  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/288334  
PRIOR FILING DATE: 2001-05-03  
PRIOR APPLICATION NUMBER: 60/291241  
PRIOR FILING DATE: 2001-05-16  
PRIOR APPLICATION NUMBER: 60/322284  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/285609  
NUMBER OF SEQ ID NOS: 155  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 75  
LENGTH: 5198  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-120-801-75

Query Match 70.9%; Score 2586; DB 14; Length 5198;  
Best Local Similarity 9.9%; Pred. No. 2.8e-29;  
Matches 423; Conservative 70; Mismatches 19; Indels 3770; Gaps 370;

QY 1 M-----SI-----SS-----D-----E-----7  
DB 1 MGRSPSWLYGVLLGALLATTCSSVNDKNDPTGKSLAFVFDITGSMFDDLVQVREGAAK 60  
QY 8 -V-----N-----F--L--VY-----13  
DB 61 IFKTVMAQREKLIYINMVPFHDPPYLGEINTDSTYFMRQLSKVYVHGGGDCPEKLTIG 120  
QY 14 -----R--Y-----L--QE--SG--F-----SHSAF-T 27  
DB 121 ILKALQISLPSSFIYFTDARSKDYLEDEVLNTIQEKSSVVFVMTGCGNRTHPGFRT 180  
QY 28 -----FG-I-----KS-----H-ISQ-----S-NI-----39  
DB 181 YEKIAAASFQVPHLEKSDVSTVLEYVRHAVKQKVHLMYAEARGGTVSRLNIPVDKHL 40  
QY 40 -----N-----GA-----LV-----P-P-----46  
DB 241 ELTISLSDKDDSDNLDIVLRDPGEKTVKRLYSKEGGTIDLKVKVLIKDKSPGVWTV 300  
QY 47 -----A-----A-----LI--S--I-----52  
DB 301 NTSRLKHTIRVFGHGAVDKFKYFASRPLDRIELARPRVNLQDTYLLINMTGLIPPGTV 360  
QY 53 --I-----Q-----KGL-----Q-Y-----VE-----AE-63  
DB 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPVFPKGLFFVRVQYDEDNFEMRIAPT 420

QY 64 -V-----S-I-----N-----ED---GTLF-----73  
DB 421 AIGSVIVGGPRAFMSPHQEFVGRDLNLSCTVESASAYTIYVWKTGEDIIIGPLFVHNTHD 480  
QY 74 -----D-----G-----R--P-I-----E-----79  
DB 481 TSVWVTIPELSLKDAGEYECRVISNNGNSVKTRETRESPPFIFGVNRVSVPLGEAFLH 540  
QY 80 -S-----L-----SL-I-----DA-----86  
DB 541 CSTRSAGEVEIRWTRYGATVFNNGPNTERTNGTLKIHVTRADAGVYECMARNAGCMST 600  
QY 87 -VM-----P-DV-----V-92  
DB 601 RKMRLDIMEPPSVKVTPODVYFNMRGVNLSCEAMGDKPVEVHWYFKGRHLLNDYKQVG 660  
QY 93 Q-----T--R--Q--QAYRD---KLA-----Q-----105  
DB 661 QDSKFLYIRDATHDEGTVECRAMSQAGQA-RDTTDLMLATPPKVEIIQNKMMVGRGDRV 719  
QY 106 -----Q-----Q-----A-----AAA-----AA-113  
DB 720 SPECTIRKPHPKIRWFKNGKDLIKPDDYIKINEGQLHIMGAKEDAGAYSCVGNMAG 779  
QY 114 -AAAA-----AS-----Q-----Q-----GS--A--KN-126  
DB 780 KDQVANLSVGRVPTLIESPHTVRVNIERQVTLQCLAVGIPPEPEIEMQKGNVLLATLNP 839  
QY 127 -GEN-----T--A--N-----132  
DB 840 RYTQADG-NLLITDAQIBDQGFCTIARNTYGOOSQSTTLMTVGLVSPVGHVPEEQ 898  
QY 133 -----G-----E-----N-----G-----137  
DB 899 IEGQDLTSLSCVVLGTPKPSIWIKDDKPVESGPTIKIEGGSLLRGNPKDEKGYTC 958  
QY 138 -A-----H-----TI-----A--N--NH-----145  
DB 959 IAVSPAGNSTLHINVQLIKKPFVYKPEGGIKPTISGMDEKHVAVVNSTHVDLDEGF 1018  
QY 146 -----T-----146  
DB 1019 AIPCVVSGTPPPIITWYLDGRPIETNSRDTVTADNTLIVRKADKSYSGVYTQATNSAG 1078  
QY 147 D-----M--M--E-----VD-----GD-----V-155  
DB 1079 DNEQKTTIRIMTPMISPGSSFNWVVDLFTIPCDVYGDPKPVITWLLDDKPFTEGVN 1138  
QY 156 E-----IP-----P-----N--K-A-----VVL 165  
DB 1139 EDGSLTIPNVNEAHRGTFTCHAQNAAGNDRVTTLTVHTTPTINAENQEKIALQNDIVL 1198  
QY 166 -----R-----G-----167  
DB 1199 ECPAKALPPPVLWTVEGEKIDSQILPHITREDGALVQNVKLENTGVFVQVSNLAGE 1258  
QY 168 -----HE-----SEV-----F--I--C-A-----WNP--V-----180  
DB 1259 SLISYTLTHHEKPKIIESEVGVVDVVKGFTEIIPCRATGVPEVIRTNKNGIDLKMEKKF 1318  
QY 181 S-D-L--L--AS--G-----SG-----D-----S--TA-----193  
DB 1319 SVDNLGLTIRIYEADKNDIGNYCVVNEAGTSQMTTHVDVQBPPIILPSTQTNNTAVGD 1378  
QY 194 RI-----W-----NL-----S-ENST-----S--204  
DB 1379 RVELKCYEASPPASVTFWFRGIAIGTDTKGYVESDGLTVIQSASVEDATIYTCASNP 1438  
QY 205 -G-----S-----TQ-----L--V-----L--211  
DB 1439 AGKAENLOVTVIASPDIKDPDVVTQESIKESHPSLYCPFSNPLPQISWYNDKPLID 1498  
QY 212 -----R--H-----CI-R-----216

Db 1499 DKTSWKTDDKRLHVFKAKITDSGVYKCVARNAAGEGSKSFQVEVIVPLNLDESQYKXK 1558  
QY 217 EG--G--Q--D--V--P--SN-----225  
Db 1559 VFAKEBEVTLGCPVSGFPVQINWVVDGTVBFGKYGATLSNGLTLHFDSVSVKQE 1618  
QY 226 KDVT-S-L-D--D--W 233  
Db 1619 GNVHCAQSGNILDIDVELSVLAVPIVGEDDNLVFLGKDISLSCDLQTESDDKTFVW 1678  
QY 234 N-SE--G--TL-L-----240  
Db 1679 SINGESDRPNVQIIPSDGRLYITDAKPENNGKYMCRVNTSAGKAERTILTLDVLEPPVF 1738  
QY 241 A--TG-----SY--D-----246  
Db 1739 VEPVFANOKLIGNPILLOQVGTGNPKPTVIWKIDGNVDKSWLFDLSLLRIEKLGT 1798  
QY 247 G--F--A--R-----I--251  
Db 1799 KSAQISCTAENKAGTASRFFIQNIAAPTFKNEGQETIPRESEITITLDCPVSLGFOIT 1858  
QY 252 W-----T-----KD-----255  
Db 1859 WMKQGLPLTENDAIFTLDNTRLTILNANRDHDIYTCVANNTAGQVSKDFVVVVQVLPKI 1918  
QY 256 CN-----L--A-----S--260  
Db 1919 KNAVTLNEGEBIILTCDAGNPPTTAKWDFNQGDLKPEAVFVNNHTVVVNNVTKYH 1978  
QY 261 L--GO--H--K-----266  
Db 1979 TGUYKYATNKVQAVKTVNHVTRTPFESGLTESELTVNLTRSTILECDVDDAIGVGI 2038  
QY 267 G-P-----I-----F-----270  
Db 2039 SWTVNGKPFLEATDGVQTLAGGRFLHVSAKTDDHGSYACTVNEAGVATKTNLFPQVP 2098  
QY 271 AL-----K--W-----274  
Db 2099 PTIVNEGBYTVIENNSLVPCEVTGKPNPVTMTKGRPVGLKSVQVLSQGPFIHV 2158  
QY 275 N-----K--KG-----278  
Db 2159 AETAHKSVCMAKNDVGTAEISFDVDIITRPMIQIKNIIVTAIKGGALPKFCPIDDDK 2218  
QY 279 NF--I-----LS-----AG-----285  
Db 2219 NFKGQIILWRNYQPIDLEAEDARITLSNDRRLTILNVTEENDEBGQYSCRVKNDAGNSFD 2278  
QY 286 V-----D--KT--TI-----I-W--D--A-H-----296  
Db 2279 FKATVLVPPTIIMLDRKNTAVEHSVTLSLCPATGKPBPDITWFKDGEAIIHENIADII 2338  
QY 297 T-----G--EA-----K-----Q 302  
Db 2339 PNGELNGNLKITRIKEGDAGKTCADNSAGSVEQDVNNVITIPKIEKGIPSDYESQ 2398  
QY 303 Q-----F--P--H-----303  
Db 2399 QNERVVISCPVYARPPAKITWLKAGKPLQSDKFVKTSAHQKLYLFLKRETDSKTYCIA 2458  
QY 304 F-----P--F-----H-----307  
Db 2459 TNEAGTKRDFKVMVLPSPFDPNPVIRITVNSGPNSTLHCPAKGSPSTITWLKDGNA 2518  
QY 308 S-----S-----A-----P-----310  
Db 2519 IEPNDRYFDFAGRQIQISKTEGSDQGYTCATNSVSGDDLENTLEVIPPVIDGERRE 2578  
QY 311 A-----L--D-----VD--WQ-----S-N-NT-----F-A--323

Db 2579 AVAVIEGFSSELFCDNSTGSDVDVEMQKQGLTINQDTLRGDSFIQIPSSGKMSFSLARKS 2638  
QY 324 S--C--S--TD-M-----S--CI-----H--332  
Db 2639 DSGRYTCIIVRNPAKEARKLDFFAVNDPPSISDLSANIQTIIVPYVPVINCVSVPSPHP 2698  
QY 333 V-----V-----CK--LG--Q--D-----339  
Db 2699 KVVWLPDDKPLEPDSAAVELTNGNETLKIVRSQVEHAGTYTCEAQNNGVKARKDLVRVT 2758  
QY 340 R-----P-----I-----KT--344  
Db 2759 APPHFEKEREVVARVGDVMTLLTCAESSVPLSVVYHHAHDESQVGVITSKYAANEKTL 2818  
QY 345 F-----F--Q-----346  
Db 2819 NVTNIQLDEGFYCTAVNEAGITKFKLVIVETPYFLDQQLYPIILGKRLTLDCSAT 2878  
QY 347 G-----H-----T--N--EV-----N-A-----354  
Db 2879 GTPPPPTILFMKDGKRLNESDEVDIIGSTLVIDNPQKEVEGRTYTCIAENKAGRSEKDMWVE 2938  
QY 355 I-----K-W-----DP-----T-----G-----N 362  
Db 2939 VLLPPKLSKEWINVEVQAGDPLTLECPIEDTSGVHITWSRQFGKQGLDMRAQSSDKSK 2998  
QY 363 LL-----A--SC-----S--D-----DMTLKI--375  
Db 2999 LYIMQATPEDADSYSCIAVNDAGGABAVQVTVNTPPKIFGDSFSFSTEIVADTTLBPCR 3058  
QY 376 W-----W-----SM--KO-----DN-----CV-----384  
Db 3059 TEGIPPEISWFLDGKPILEMPGVYKQGLSLRIDNIKPNQEGRTYTCVAENKAGRAEQD 3118  
QY 385 -----384  
Db 3119 TVVEISEPPRVVMASEVMRVEGRQTTIRCEVFNPEPVVNMKGEPYVTDLLQFSTKL 3178  
QY 385 H-----D--L-----Q-----O-----H--390  
Db 3179 SYLHRETTLAGGTYTCIATNKAQESQTTDVEVLVPPRIEDSERVLOGKEGNTVMVHC 3238  
QY 391 N--KEI-----Y-----395  
Db 3239 QVTGRPVVYVTKRNGKEIEQFNPVLHINATRADEGKTSIASNEAGTAVADFLIDVFT 3298  
QY 396 T-----T-----I--K--W-----SP-----401  
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISMLKGGRPFNMDNIILSPRGDTLMILK 3358  
QY 402 -----T-----G--403  
Db 3359 AORPDGLYTCVATNSYGDSEQDFKVNVTYKPIDETIDQTPKANAGSIIILKCPVLGNP 3418  
QY 404 P-----G-----T--NN-----408  
Db 3419 TTVTWKRGDDAVPNDPSRHTIYNNVDLKNSTVEDAGQYSCIAVNEAGNLTHYAAEVI 3478  
QY 409 P-----N--AN--L-M-----LASAS-F-----420  
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYDRYSISPDGSHIT 3538  
QY 421 D-----D-----S--T-----V-R--L-----426  
Db 3539 INKAKLSGKGYICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIYECPI 3598  
QY 427 W-----W-----DV--D--R-----GI-----CI-----H--436  
Db 3599 SGIPQPDVUWTKNGMDINNTDSRVILAQNNFTFGIENVQVTDQGYTCATNRGKASHD 3658  
QY 437 TLT--K--H--Q-----E-----P-----444  
Db 3659 FSLDVLSPPEFDHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALODLT 3718

```
QY 445 --V--YS-VA-----F-----SP-----D-----453
Db 3719 DNVDISDDGKLTISOASLENAGLYTCIALNRAGEASLEFKVEILSPFVIDSRNDVQ PQ 3778
QY 454 -----G-----R-----Y-L-AS--G--S-461
Db 3779 VAVNQPTIMRCAVTGHPFPISKWLNKQKEVTDENIRIVEQGQVLQILRTDSDHACKWSC 3838
QY 462 -----F-----D--K-----C-----465
Db 3839 VAENDAGVKELEMDVFTPPVSVKSDNPIKALGETITLFCNAGSNPPQLKWAAGGSL 3898
QY 466 -----V-H-----I-----468
Db 3899 IFDSPDGAISLKGARLDIPLHKKTDVGYTCOALNAAGTSEASVSDVLVPPEINRDI 3958
QY 469 -----W-N-----T-Q-----472
Db 3959 DMSPLPAQQLSLTQCLAQKPVQMRWTLNGTALHSTPGITVASDSTFIQINNVSLS 4018
QY 473 --V--C-----LHY-----L-NG--QV-----L-484
Db 4019 KGYTCTAENAVAGSDNLMYNDVYQAPVISNGGTQKQVIEGELAVIECLVEGYPAQVSWL 4078
QY 485 --N-----L--GR-----S-I--C-----LVTL-----495
Db 4079 RGNRVETGVQVRYVTDGRMLTIIEARSLDSGIYLCSATNEAGSAQAYTLEVLSPKI 4138
QY 496 -----P-----H-----497
Db 4139 ITSTPGVLTPSSGSKFSLCAVRGYDPPIISWTLNGNDIKDNGHTIGADGTLHIEKAE 4198
QY 498 --HLV-----V--I--P-----LV-----AL--IE-----509
Db 4199 ERHLIYECTAKNDAGDTLEFPVQTVAPKISTSGNRYNGSEGTETVIKCEIESSEB 4258
QY 510 -----L--LV-----LK 514
Db 4259 SWSKNGVPLLPNNLIFSEDDYK 4280
```

```
RESULT 15
US-10-369-493-6858
; Sequence 6858, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 6858
; LENGTH: 5198
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
US-10-369-493-6858
```

```
Query Match 70.9%; Score 2586; DB 14; Length 5198;
Best Local Similarity 9.9%; Pred. No. 2.8e-29;
Matches 423; Conservative 70; Mismatches 19; Indels 3770; Gaps 370;
QY 1 M-----SI-----SS-----D-----E-----7
Db 1 MGRSPSWLYGVGLGILLIATTCSSVNDKNDPTGKSSLAFFVFDITGSMFDDLQVREGAAK 60
```

```
QY 8 --V--N-----F--L--VY-----13
Db 61 IFKTWAQREKLIYNYIMVPHDPYLGEIINTDSTYFMRQLSKVYVHGGDCPEKTLTG 120
QY 14 -----R--Y-----L--OE--SG--F-----SHAP--T 27
Db 121 ILKALQISLPSSFYVFTDARSKDYHLEDEVLTITQEKQSSVVFVMTGDCGNRTHRGFT 180
QY 28 -----RG-I--KS-----H-ISO-----S-NI-----39
Db 181 YEKIAAASFGVFHLEKSDVSTVLEYVRHAVKQKKVHLMYEARERGGTVSRNIPVDKHL 240
QY 40 -----N-----GA-----LV-----P-P-----46
Db 241 ELTISLSDGKDDSDNLDIVLRDPEGTVDKRLYSKEGGTIDLKNVKLIRLKPSPGVTV 300
QY 47 -----A-----A-----LI--S--I-----52
Db 301 NTNSRLKHTIRVFGHGAVDKYGFAFRPLDRIELARPRVLNQDTYLLINMTGLIPPGTV 360
QY 53 --I-----Q-----KGI-----Q-Y-----VE--AE--63
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYPAGFPVPPKGLFFVRVQSDYEDNEFEFRIAPT 420
QY 64 --V--S--I-----N-----ED--GTLF-----73
Db 421 AIGSVLVGFPRAFMSPIHQEFVGRDLNLSCTVESASAYTIYVWKTGEDIIIGPLFYHNTD 480
QY 74 -----D-----G-----R--P--I-----E-----79
Db 481 TSVWTIPELSLKDAGEYECRVISNNGNYSVKTRVETRESPPEIFGVNRVNSVPLGEAFLH 540
QY 80 --S-----L-----SL-I-----DA-----86
Db 541 CSTRSAGEVEIRWTRYGATVFNQPNTERNPTNGTLKHHVTRADAGVCECMARNAGMST 600
QY 87 --VM-----P-DV-----V-----92
Db 601 RKMRLDIMEPPSVKVTQPDVYFNMREGVNLSCAMGDPKPEVHWYFKGRHLLNDYKQVG 660
QY 93 Q-----T--R--Q--QAYRD-----KLA-----Q-----105
Db 661 QDSKFLYIRDATHDEGTCECRAMSQAQA-RDITDLMLATPPKVEIIONKMMVGRGDRV 719
QY 106 -----Q-----Q-----A-----AAA-----AA--113
Db 720 SPECKTIRKPHPKIRWFKNGKDLIKPDYIKINEGLHMGAKDEDAAGAYSCVGENMAG 779
QY 114 --AAAA-----AS-----Q-----Q-----GS--A--KN--126
Db 780 KDQVANLSVGRVPTTIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNP 839
QY 127 -----GEN-----T--A--N-----132
Db 840 RVTQLADG-NLLITDAQIEDQGTCTIARNTYQOSQSTLLMVTGLVSPVLGHVPPEEQ 898
QY 133 -----G-----EE-----N--G-----137
Db 899 IEGQDLTSLCVVLTGTPKPSIWMKDDKPEVEGPTIKIEGGSLRLRGNPKDEKGYTC 958
QY 138 -----A-----H-----TI-----A--N--NH-----145
Db 959 IAVSPAGNSTLHNVQLIKKPEFVYKPEGGIVFKPTISGMDEKHVAVNSTHVDLGE 1018
QY 146 -----T-----146
Db 1019 AIPCVVSGTTPPIITWYLDGRPIITNSRDTVTADNTLIVRKADKSYSGVYTQATNSAG 1078
QY 147 D-----M--M--E-----VD-----GD-----V--155
Db 1079 DNEQKTTIRMTNTPMISPGQSSFNMMVVDLFTPCDVGDPKPVITWLLDDKPFTEGVN 1138
```

QY 156 E-----IP-----P-----N-K-A-----VVL 165  
DB 1139 EDGSLTIPVNEAHRGFTTCHQAQAAGNDRTVTLTVHTTPTNAENOEKIALQNDIVL 1198  
QY 166 -----R-----G-- 167  
DB 1199 ECPAKALPPVRLWYVEGEKIDSQLPHITIREGALVQNVKLENTGVFCVQVSNLAGE 1258  
QY 168 -----HE-----F-----I-C-A-----WNP-V----- 180  
DB 1259 SLSYTLTVHEKPKLISEVPGVVDVVGFTTIEIPCRATGVPEVIRTNKNGIDUKMEKFF 1318  
QY 181 S-D-L-L-----G-----SG-----D-----S-----TA----- 193  
DB 1319 SVDNLGTLRIYADKNDIGNYCNVWNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVGD 1378  
QY 194 RI-----W-----NL-----S-ENST-----S-- 204  
DB 1379 RVELKCYVEASPPASVTWFRFGIAIGTDTKGYVVESDGTILVIQASVEDATIYCKASNP 1438  
QY 205 -G-----S-----TO-----L-V-----L-- 211  
DB 1439 AGKAELANQVTIASPDIKDPDVVTQESIKESHPPSLYCPVFSNPLPQISWYLNDRPLID 1498  
QY 212 -----R-H-----CI-R----- 216  
DB 1499 DKTSWKTSDDKRKLHVFKAKITDSGVYKCVARNAAGBSKSFQEVIVPLNLDESXYKKK 1558  
QY 217 -EG-----G-----Q-----D-----V-P-----SN----- 225  
DB 1559 VFAKEGEVTLGCPVSGFPVQINWVVDGTVVFBPKKYGATLSNDGLTLHFDSVSVKOE 1618  
QY 226 -----KVT-S-L-----D-----W 233  
DB 1619 GNHVCVAQSGKNILIDVLSVLAVPIVGEDDNLVFLGDKDISLSCDLQTESDKTTFVW 1678  
QY 234 -N-SE-----G-----TL-L----- 240  
DB 1679 SINGESDRPDNVQIPSDGHRLYITDAKENNKKYMCRTNSAGKAERTLTLDVLEPPVF 1738  
QY 241 -----A-----TG-----SY-D----- 246  
DB 1739 VEPVFEANOKLIGNPILLOQVGTGNPKPTVIWKIDGNDVDSKSWLFDLSLLRIEKLGT 1798  
QY 247 -----G-----F-----A-----R-----I- 251  
DB 1799 KSAQISCTAENKAGTASRDPFQIONIAAPTKEGQDFTIFRESEITLDCPVSLGDFQIT 1858  
QY 252 W-----T-----KD----- 255  
DB 1859 WMKQGLPLTENDALFTLDNTRLTLNANRDHDIYTCVANNTAGQVSKDFVQVVLPKI 1918  
QY 256 -----CN-----L-A-----S----- 260  
DB 1919 KNAVVTLEINEGEIEILTCAEGNPTPTAKWDPNQGDLPKEAFVANNHTVVVNNVTKYH 1978  
QY 261 T-----L-----CO-----H-K----- 266  
DB 1979 TGVIKCYATNKVGQAVKTNVHVTRKPRFESGLTESELVNLTRSLTECDVDDAIGVGI 2038  
QY 267 -----G-P-----I-----F----- 270  
DB 2039 SWTVNGKPLAETDGVQTLAGGRPLHVSAKTDDHGSYACTVNEAGVATKTNLFVQVP 2098  
QY 271 -----AL-----K-----W----- 274  
DB 2099 PTIVNEGGEVTVIENNSLVLPCVETCKPNPVVTTKDGPRVGLKSVQVLSQGFQKIVH 2158  
QY 275 -----N-----K-----KG----- 278  
DB 2159 AEIAHGSYICMAKNDVGTAEISFDVDIITRPRIQRIKNIIVTAIRGGALPFXCPIDDDK 2218  
QY 279 NF-----I-----LS-----AG----- 285

DB 2219 NFKGQIILWRNYQPIDLEAEADARITRLSNDRRRLTLNVTNTEDEGQYSCRKNDAGENSFD 2278  
QY 286 -----V-----D-KT-----TI-----I-W-D-A-H----- 296  
DB 2279 FKATVLVPPPTIIMLDKDKNKTAVEHSTVTLSCPATGKPEPDITWFKDGEAIIHENIADI 2338  
QY 297 -----T-----G-----EA-----K-----Q 302  
DB 2339 PNGELNGNLKTRIKEGDAGKYTCADNSAGSVEQDVNVNVTIPKIEKDGPSPDYEQ 2398  
QY 303 Q----- 303  
DB 2399 QNERVVISCPVAVPPAKITWLGKAPLQSKDFVKTSAHQKLYFLKRETDSSKYTCIA 2458  
QY 304 -----F-----P-F-----H----- 307  
DB 2459 TNEAGTKRDFKVMVLPVAFSDFEPNIVRRITVNSGNPSTLHCPAKGSPSPITWLDKQNA 2518  
QY 308 -----S-----A-----P----- 310  
DB 2519 IEPNDRYVFDAGRQIQISKTEGSDOGRYTCIATNSVSGDDLENTLEVIIPPIDGERRE 2578  
QY 311 -A-----L-D-----VD-WQ-----S-N-NT-----F-A----- 323  
DB 2579 AVAVIEGFSSELFCDSNSTGVDVWQDGLTINQDTRLGDSFTQIPSSGKKMSFLSARKS 2638  
QY 324 -S-----C-----S-TD-M-----CI-----H- 332  
DB 2639 DSGRYTCIVRNPAGEARKLFDFAVNDPPPSIDELSSANITQIVPYVPVINCINVSVP 2698  
QY 333 -----V-----CK-----LG-O-D----- 339  
DB 2699 KYWLFDDKPLBPSAAYELTNNGETLKIIVRSQVEHAGTYTCEAONNVGKARKDFLRVT 2758  
QY 340 -----R-----P-----I-----KT- 344  
DB 2759 APHPKEEREVEVVARVGDWMLTCAESSVPLSVVYHHAHDESVQNGVITSKYAANEKTL 2818  
QY 345 -----F-O----- 346  
DB 2819 NVTNIQLDDEGFYCYTAVNEAGITKFKPLIVETPYFLDQOKLYPIILGKRLTDCSAT 2878  
QY 347 G-----H-----T-N-----EV-----N-A----- 354  
DB 2879 GTPPPITLFWKDKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGRSEKMMVE 2938  
QY 355 I-----K-W-----DP-----T-----G-----N 362  
DB 2939 VLLPPKLSKEWINVEVQAGDPLTLECPIDTSGVHITWSRQCGKQGLDMRAQSSSDSKS 2998  
QY 363 LL-----A-SC-----S-----D-----DMTLKI----- 375  
DB 2999 LXIMOATPEDADSYSCIAVNDAGBAFQVTVNTPPKIFGDSFSTTEIVADTTLTLEPCR 3058  
QY 376 -----W-----SM-----KO-----DN-----CV----- 384  
DB 3059 TEGIPPPBISWFLDGPILMPGVYTKQGLSLRIDNIKNPQEGRYTCVAENKAGRAEQD 3118  
QY 385 ----- 384  
DB 3119 TYVEISEPPRVVMASEVMRVVEGRQTTIRCEVFGNPEPVVNLKDGEPYTSLLQFSTKL 3178  
QY 385 H-----D-L-----Q-----O-----H- 390  
DB 3179 SVLHRETTLADGGTYTCIATNKAGESOTTDDVEVLVPPRIEDEERVLOGKEGNTYMHVC 3238  
QY 391 -----N-KEI-----Y----- 395  
DB 3239 QVTGRPVVVTWKRNGKEIEQFNVLHIRNATRADEGKYSCIASNAGTAVADFLIDVFT 3298  
QY 396 -----T-----I-----K-----W-----SP----- 401

Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRPFNMNDNIILSPRGDTLMILK 3358  
QY 402 -----T-----G-- 403  
Db 3359 AQRFDGGLYTCVATNSYGDSEQDFKVNVTYKPYIDETIDQTPRAVAGEIILKCPVLGNP 3418  
QY 404 -P-----G-----T--NN----- 408  
Db 3419 TPVTWKRGDVAVPNSRHTIVNNYDLKINSVTTEDAGQYSCIADVNEAGNLTHYAAEVI 3478  
QY 409 --P-----N--L-M-----LASAS-F----- 420  
Db 3479 GKPTFVRKGNLVEIENDTITWDCGVTSRPLFSISWFRGDKPVLYLDVYSISPDGSHIT 3538  
QY 421 -----D-----S--T-----V-R--L----- 426  
Db 3539 INKAKLSGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPI 3598  
QY 427 -----W-----DV--D-R-----GI-----CI-----H- 436  
Db 3599 SGIPQPDVITWKNGBDINMTDSRVILAQNNETFGIENVQVTDQGRYTCATNREGKASHD 3658  
QY 437 ----TLT--K--H--Q-----E-----P----- 444  
Db 3659 FSLDVLSPPEFDIHGTQPTIKREGDTITLTCPILAEADIAQVMDVSWTKDSRALDGLT 3718  
QY 445 --V-----YS-VA-----F-----SP-----D----- 453  
Db 3719 DNVDISDGRKLTISOASLENAGLYTCIALNRAGEASLEFKVEILSPVVIDISRNDVQFQ 3778  
QY 454 -----G-----R-----Y-L--AS--G--S- 461  
Db 3779 VAVNQPTIMRCVTHGPPFSIKWLKNGKEVTDENIRIVEQGVQLILRTDSDHAGKWSC 3838  
QY 462 -----F-----D--K--C----- 465  
Db 3839 VAENDAGVKELEMLVDFTPPVSVKSDNFIKALGETITLFCNASGNPYPOLKWKAGGSL 3898  
QY 466 -----V-H-----I 468  
Db 3899 IPDSPGARISLKGARLDIPLHKKTDVGDYTCQALNAAGTSEASVSDVLVPPPEINRDGI 3958  
QY 469 -----W--N-----T--Q----- 472  
Db 3959 DMSPRLPAAQSLTLOCLAQKPVQMRWTLNGTALTHTSPGITVASDSTFIQINNVSLS 4018  
QY 473 --V-C-----LHY-----L-NG--QV-----L 484  
Db 4019 KGVYTCVAENVAGSDNLMYVNDVVQAPVISNGGKQVIEGELAVIECLVEGYPAPQVSWL 4078  
QY 485 ---N-----L--GR-----S-I-C-----LYTL----- 495  
Db 4079 RNGNRVETGVQVRYVYTDGRMLTIIEARSLDSGILCSATNEAGSAQAYTLEVLVSPKI 4138  
QY 496 -----P-----H----- 497  
Db 4139 ITSTPGVLTPSSGSKFSLPCAARGYPDPITISWTLNGNDIKDGENGTIGADGTLHIEKAE 4198  
QY 498 --HLV-----V--I--P-----LV-----AL---IE----- 509  
Db 4199 ERHLIVECTAKNDAGADTLEFPVQTIIVAPKISTSGNRYNGSEGTETVIKCEIESESSEF 4258  
QY 510 -----L-----LV-----LK 514  
Db 4259 SNSKNGVPLLPNSNLIFFSEDIK 4280

Search completed: January 3, 2005, 15:48:12  
Job time : 137 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:09:41 ; Search time 21.6667 Seconds  
(without alignments)  
2282.558 Million cell updates/sec

Title: US-09-987-701-4

Perfect score: 3649

Sequence: 1 MSISDEVNFLVRYLQESG.....LPHLVVIPLVALIELLVLK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2589.3	71.0	7962	2 I38346	elastic titin - hu
2	2586	70.9	5175	2 T20992	hypothetical prote
3	2586	70.9	5198	2 T43290	hemacentin precurs
4	2585	70.8	10797	2 T30192	probable peptide s
5	2575.8	70.6	26926	1 I38344	titin, cardiac mus
6	2575.1	70.6	4660	2 T42737	gp330 protein prec
7	2574.2	70.5	7829	2 T15789	hypothetical prote
8	2572.1	70.5	6658	2 T13931	projectin - fruit
9	2568.8	70.4	5232	2 A45086	HC-toxin synthetas
10	2568.3	70.4	3461	2 S58870	reelin precursor -
11	2566.1	70.3	5825	2 T12117	polyprotein - fava
12	2562.6	70.2	4544	1 S02392	alpha-2-macroglobu
13	2556	70.0	4543	1 A53102	alpha-2-macroglobu
14	2553.7	70.0	4753	1 A47437	LbL-receptor-relat
15	2552.7	70.0	5376	2 T42215	zonadhesin - mouse
16	2551.4	69.9	4545	1 S25111	alpha-2-macroglobu
17	2550.6	69.9	6831	2 A88852	protein unc-22 (im
18	2550.6	69.9	6839	2 S57242	twitchin (similar
19	2550.6	69.9	7160	2 T27935	hypothetical prote
20	2549.2	69.9	8243	2 T31307	type I fatty acid
21	2549.1	69.9	6805	2 S20901	titin - rabbit (fr
22	2548.4	69.8	6669	2 S55024	nebulin, skeletal
23	2537.3	69.5	9376	2 T14593	syringomycin synth
24	2533.4	69.4	15281	2 S41309	cytosporin synth
25	2528.8	69.3	3623	2 T09456	intrinsic factor-B
26	2527.7	69.3	4861	2 S71752	giant protein p619
27	2527.3	69.3	4767	2 T31345	hypothetical prote
28	2527.1	69.3	3623	2 T08618	intrinsic factor-B
29	2527	69.3	4836	2 T14346	herc2 protein - mo

30	2523	69.1	5369	2 T44807	myosubtilin synth
31	2516.8	69.0	4464	2 D87755	protein T21E12.4 [
32	2516.6	69.0	5138	2 B96695	hypothetical prote
33	2515.4	68.9	4302	2 A38971	polycystic kidney
34	2514.8	68.9	4930	2 E69679	polyketide synthet
35	2514.5	68.9	3871	2 T22812	hypothetical prote
36	2513.4	68.9	4568	2 T08030	dynein beta heavy
37	2511.8	68.8	4572	2 S57908	hypothetical 537K
38	2502.3	68.6	4924	2 T50176	probable peptide s
39	2501.8	68.6	6359	2 T31679	bacitracin synthet
40	2500.9	68.5	4644	1 A38905	dynein heavy chain
41	2500.5	68.5	4447	2 A69679	polyketide synthas
42	2498.9	68.5	4639	1 A54794	dynein heavy chain
43	2498.8	68.5	6642	2 T29757	protein UNC-89 - C
44	2496.5	68.4	6486	2 T31076	tyrocidine synthet
45	2496	68.4	3856	2 T51174	ataxia-telangiecta

ALIGNMENTS

RESULT 1

I38346

elastic titin - human (fragment)

C:Species: Homo sapiens (man)

C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 09-Jul-2004

C:Accession: I38346

R:Labett, S.; Kolmerer, B.

Science 270, 293-296, 1995

A:Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.

A:Reference number: A57430; MUID:96026330; PMID:7569978

A:Accession: I38346

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-7962 <RES>

A:Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PIDN:CAA62189.1; PID:g101-

C:Genetics:

A:Gene: GDB:TTN

A:Cross-references: GDB:127867; OMIM:188840

A:Map position: 2q31-2q31

Query Match

Best Local Similarity 71.0%; Score 2589.3; DB 2; Length 7962;

Matches 454; Conservative 50; Mismatches 8; Indels 4667; Gaps 390;

QY	1	MS-----IS-----SDE-----V-N-----	-----9
DB	57	MTQFEDTYQLEIAEAYPEDEGTYTFVANNVQVSVSTANLSLEAPESILHERIEQIEIME	116
QY	10	-----F-----L-----VY-----R--Y--15	
DB	117	MKAAPVKKRBPLEVALGHLAKFTCEIQSAPNVRPQWFRAGREIYESDKSIRSSKIYS	176
QY	16	----L--Q-----E-----S-----G---F--21	
DB	177	SLEILRTQVDCGEYTCRASVEYGVSCATLTIVTEAYPTFLSRPKSLTTFVGKAKFI	236
QY	22	-----S-----H-----	24
DB	237	CTVTGTPVETIWKDGAALSPNWRISDAENKHLELSNLTIDRGVYVSCASNKFGA	296
QY	25	----A-----F-----T-----	27
DB	297	DTCQAEIIIDKPHFIKLEPVSQAINKKVHLEQCVDKRVTVTWKDGQKLPPGKDYK	356
QY	28	--F-----G-----IK-----	31
DB	357	ICFEDIATLEIPLAKDKSGTVVCTASNEAGSSCSATVTVREPPSFVKYVDPSPYLMLP	416
QY	32	--S---H-----ISQ-----SN-----IN-----G-----41	
DB	417	GESARLHCKLKGSPVI-QVTWFKNNKELSESNTVRMYFVNSEAILDITDVKVEDSGSYSC	475

QY 42 -A- -LV- -P- -PA- -AL- -IS- 51  
Db 476 EAVNDVSGSDSCSTEIVIKPBPPIKLEPADIVRGTNALLQCEVSGTGPFPEISWFKDKQ 535  
QY 52 -I- -I- -I- -I- -I- 52  
Db 536 IRSSKKYRLFQSKLSVLEIFPSNSADVGEYECVAVNEVGKCGMATHLLKBPPTFVKV 595  
QY 53 -I- -Q- -KG- -L- -Q- 58  
Db 596 DDLIALGQGVTLQAARVSEPISTVMKQGEVIREDKIKMSFNGVAVLIPDVQISF 655  
QY 59 -Y- -VE-AE- -I- -I- -I- 63  
Db 656 GGYTCLAENAGSQTSGELIVKEPAKIERAELIQVTAGDPATLEYTVAGTPELKPXW 715  
QY 64 -V-S- -I- -N- -E-D- -G- -T-L- 72  
Db 716 YKGRPLVASKKYRISFKNNVAQLKPYSAELHDSGOYTFEISNEVSGSCETTFTVLDRD 775  
QY 73 -F- -D- -G-R- -P- -I- -E-SL 81  
Db 776 IAPFFTPLRNVDSVVNGTCLDCKIAGSLPMRVSWFKDGEIAASDRYRIAFVEGTASL 835  
QY 82 SLI- -DA- -V-MP- -DV- -V- -Q 93  
Db 836 EIIRVMDNAGNFTCRATNSVGSKDSSGALIVQEPSPFTKPGSKDVLPGSAVCLKSTFQ 895  
QY 94 -T-R- -I- -I- -I- -I- -I- 95  
Db 896 GSTPLTIRFKGNKELVSGSCVITKEALESLELYLVKTSDSGTCTCKVSNVAGGVECS 955  
QY 96 -Q- -Q- -A- -Y- -I- -I- 99  
Db 956 ANLFVKEPATFVEKLEPFSOLLKGDATOLACKVTGTPPIKITWFANDREIKESSKRMSP 1015  
QY 100 -R- -D- -D- -D- -K- -K- 102  
Db 1016 VESTAVLRLLTDVIGESGEYMCQAQNEAGSDHCSSIVIVKESPYFTKEPKEIYVLEKDYV 1075  
QY 103 -LA- -Q- -Q- -Q- -Q- -Q- 107  
Db 1076 MLLAEVAGTPPPEITWFKDNTILRSGRKYKTFIQDHLVSLQLKFAADAGEYQCRVTNE 1135  
QY 108 -A- -A- -A- -A- -A- 111  
Db 1136 VGSSICSAVTLREPPSPFKKIBESTSLRGGTAAFOATLKGSLPITVTWMLKDSDEITDD 1195  
QY 112 -A- -A- -A- -A- -A- -AAS- 119  
Db 1196 NIRMFTFENNVAISLYLSGIEVKHDGKYVCOAKNDAGIQRCALLSVKEPATITEEAVSIDV 1255  
QY 120 -Q- -G- -SAK- -I- -I- -I- 125  
Db 1256 TQGDPATLQVFGSGTKETAKWFKDQGLTLGSKYISVTDVTSILKIISTEKKKSGEYV 1315  
QY 126 -N- -G- -G- -G- -G- -G- 127  
Db 1316 FEVQNDVGRSSCKARINVLDLIIPSPFTKLLKMDSIKGSFIDLECIVAGSHPIQWFK 1375  
QY 128 -E- -NTA- -N- -G- -E- -E- 134  
Db 1376 DDOEISASEKYSFHDNTAFLEISQLEGTDSGTTCATNKAGHNQCSGHLTVKEPPYF 1435  
QY 135 -E- -N- -G-A- -H- -H- -H- 139  
Db 1436 VEKPSQDVNPNTRVOLKALVGGTAPMTIKWFKDNKELHSGAARSVMKDDTSTSLFAA 1495  
QY 140 -T-I- -A- -A- -N- -N- -N- 143  
Db 1496 KATDSGTIYICQLSNDVGTATSKATLFVKEPPQPIKPPSPVLVRNGOSTTFECQITGTPK 1555  
QY 144 -N- -H- -T- -D- -D- -D- 147

Db 1556 IRVSWLDGNETAITQKHGISFIDGLATFOISGARVENSCTVVCARNDAGTASCSELK 1615  
QY 148 -MM- -EV- -D- -G- -D- -V- 155  
Db 1616 VKEPPTFIRELQPVVKVYSDVELECEVTGTPTPPFTWLKNNREIRSSKKYTLTDRVSFV 1675  
QY 156 -E- -I- -PP- -N- -K- -K- 161  
Db 1676 NUHITKCDPSDTGEYQCIUSNEGSCSTRVALKEPFSFIKIENTTVLKSSATFQST 1735  
QY 162 -A- -V- -L- -R- -GH- -PS- 170  
Db 1736 VAGSPISITWLKDDQILDDEDDNVVISFVDSVATQIRSVDNHSGRYCTOAKNESGVER 1795  
QY 171 -I- -I- -I- -I- -I- -I- 170  
Db 1796 CYAFLLVQEPQAIQVEKAKSDVDTEKDPMTLECVVAGTPELKYKWLKDGQIIVPSRYFMS 1855  
QY 171 -E- -E- -V- -P- -F- -F- 173  
Db 1856 FENNVAIFQSVWKQDSGOYTFKVENDFGSSSCDAYLVLQNDIPPSFTKLTWMDKVL 1915  
QY 174 -I- -C- -A- -W- -I- -I- 177  
Db 1916 GSSIHMCKVSGSLPISAQWPKDGKEISTSAKYRLVCHERSVSVLEVNLEDEDTANYTCK 1975  
QY 178 -N- -P- -V- -S- -D- -D- 182  
Db 1976 VSNVAGDDACSGILTVKEPPSFLVKPGRQQAIPDSTVEFKAILKGTTPPKIKWFKDDVEL 2035  
QY 183 -LL- -AS- -GS- -GS- -GS- 188  
Db 2036 VSGPKCFGLESTFSLNLYSVDASKTGOYTCHVTNDVSGSDCTTMLLVTEPPKFKVKKLE 2095  
QY 189 -GDST- -A- -R- -IW- -N- -L- -S- -E- -N- -ST- 203  
Db 2096 ASKIVKAGDSRSLCKEAGSPEIRVWVFRNEHPSALDKYRMTFIDSAVAVIQMNNLISTED 2155  
QY 204 SG- -ST- -I- -I- -I- -I- -I- -V- 210  
Db 2156 SGDFICEAQNPAQSTSCSTKVIKPPVPFPPPIVETLKNAEVSLCSELSGTPPEPVW 2215  
QY 211 -LR- -HC- -HC- -HC- -HC- 214  
Db 2216 YDKQRLSSKKYKIASKNFHTSIHLNVDTSDIGYHCKAQNEVGSCTVCTVKLKEPP 2275  
QY 215 -IRE- -IRE- -IRE- -IRE- -IRE- 217  
Db 2276 RFVSKLNSLTVVAGEPAELQASIEGAQPIFVQWLKEKEVIRESENIRITFVENVATLOF 2335  
QY 218 -GG- -GG- -GG- -GG- -GG- 219  
Db 2336 AXAEPANAGKYTCQIKNDGGMENMATLMLVLEPAVIVEKAGPMTVTVGSTCTLECKVAGT 2395  
QY 220 -QD- -QD- -QD- -QD- -QD- 221  
Db 2396 PELSVENYKDKLLTSSQKHESFYNNKISSLLSVRQDAGTYTTFVQVQNVGKSCTAV 2455  
QY 222 -V-PS- -N- -K- -K- -K- 226  
Db 2456 VDVSRAVPPSPTRRLKNTGGVLGASCILECKVAGSSPISVAWFHEKTKIVSGAKYQTTF 2515  
QY 227 -D-V-T- -SLD- -I- -I- -I- 232  
Db 2516 SDNVCTQLNSLSDSDMGNYTCVAANVAGSDECRALVTVOEPPSPFVKEPELEVLPGKNV 2575  
QY 233 -W- -W- -N- -S- -E- -E- 236  
Db 2576 TFTSVIRGTPPFKNVWFRGARELVKDCRCNIYFEDTVAEELFNIDISQSGEYTCVVSNN 2635  
QY 237 -G- -T-LL- -A- -TG- -S- -Y- -DGF- 248

Db 2636 AGQASCTTRLFVKKEPAFLKRLSDHSEVPEGKSIILESTYTGTLPISTVTKWKGDNITTSE 2695  
QY 249 -----A-R----- 250  
Db 2696 KCONIVTEKTCILEILNSTKRDAGQYSCBIENAGRDVCGALVSTLEPPYFVTELESPLEA 2755  
QY 251 -----I-W-TK-----D-G--- 256  
Db 2756 AVGDSVSLQCVAGTPEITVSMYKGTGKLRPTPEYRTYFTNNVATLVFNKNVINDSGEYT 2815  
QY 257 -----N-L-AS-----TL-----G-Q--- 264  
Db 2816 CKAENSIGTASSKTVFRIQERQLPPSFARQLKDIEQTVGLPVLTCRLNGSAPIQVCWYR 2875  
QY 265 -----H-----K-----G----- 267  
Db 2876 DGVLLRDHENLQTSFVDNVATLKILQDLSHSGQYSCSASNPLGTASSNARLTAREPKS 2935  
QY 268 -----P-----I-----F-----A-LK-W-NK----- 276  
Db 2936 PFDIKPVSIDVIAGESADFECHVTGAQPMRITWSKDNKEIRPGNGYTITCVGNTPEHLRI 2995  
QY 277 -----KG----- 278  
Db 2996 LKVGKDSGOYTQOATNDVKOMCSAQLSVKEPPKFVKLEASKVAKQGESIQLECKISG 3055  
QY 279 -----N-FI-----L-----SA-----GV-D----- 287  
Db 3056 SPEIKVSWFRNSELHESKYNMSPFINSVALLTINEASAEEDSGDYICEAHNGVGDASCST 3115  
QY 288 -----K-----T-----T-----TW----- 293  
Db 3116 ALTVKAPPVFTQKPSVPGALKSGDVILOCEISGTPPFVFWVKDRQVRNKKFKITSKH 3175  
QY 294 -----A-----H-----T-----G-R- 299  
Db 3176 FDTNLHILNEASDVGEYHCKATNEVSGDTCSCSVKFKEPFRPVFKLSDTSTLIGDAVEL 3235  
QY 300 -----A-----K-----Q----- 303  
Db 3236 RAIVEGQFQISVVLWKORGEVIRESENTRISFIDNIATLQLSPEASNSGKICQIKNDA 3295  
QY 304 -----F-----P----- 306  
Db 3296 GMRECSAVLTLEPARIEKEPEMTVTGPNFALECVVGTGPPELSAKWFKDGRSLSADSK 3355  
QY 307 -----H-S-----A-----P-A-----L-----D-V----- 315  
Db 3356 HHITFINKVASLKIPCAEMSDKGLYSFEVKNSVGKSNCTSVSHVSDRIVPPSFIRKLKDV 3415  
QY 316 -----W-----QS-----N----- 319  
Db 3416 NAILGASVVLECRVGSAPISVGFQDNEIVSGPKQCSFSFSENVCTNLNLSLEPSDTGI 3475  
QY 320 -----N-----TF----- 322  
Db 3476 YTCVAANVAGDECSAVLTQBPSPPEQTPDSVEVLPNGSLFTFTSVIRGTPPPKWKFKG 3535  
QY 323 -----ASC----- 325  
Db 3536 SRELVPESCNISLEDFVTELELFEVQPLESGDYSCLVTNDAGSASCTTHLFVKEPATFV 3595  
QY 326 -----S-----T-----D-----M----- 329  
Db 3596 KRLADPSVETGSPVLEATYTGTPPISVSWIKDEYLIOSERCISITWTEKSTILELEST 3655  
QY 330 -----C-----I-----HV-----CKL-G----- 337  
Db 3656 IEDYAQYSCLIENEAGQDICEALVSLVLEPPYFIEPLEHVEAVIGEBPATLQCKVDGTPEIR 3715  
QY 338 -----Q-----D----- 339  
Db 3716 ISWYKEHTKLRSAKYKMQFKNNVASLVINKVDHSDVGEYSCKADNSVGAVASSAVLVIK 3775

QY 340 -----R-----P-----I-KT-----F-----Q-----G-----H--- 348  
Db 3776 ARKLPFPFARKLKVHETLGGPPVAFECRINGSEPLQVSMYKGVLLKDDANLQTSFVHNV 3935  
QY 349 -----T-----N-----EV----- 352  
Db 3836 ATLQILQTDQSHIGQYNCASNPLGTASSAKLILSEHEVPPFDLKPVSVDLALGESSGT 3895  
QY 353 -----NA-IK-W-D-----PTGN-----L-L-----A-SC-S--- 368  
Db 3896 FKCHVTGTAPIKITWAKONREIRPGNYKMTLVENTATLTVLKVGKGDAGQYTCYASNIA 3955  
QY 369 -----D-----DM-T-L-KI-----W-----SMK-Q--- 380  
Db 3956 GKDSCSAQLGVQVQPPFIKKLPSRIVKQDEFTRYECKIGSGPEIKVLWYKDET-EIQES 4014  
QY 381 -----D-----N-----C-----V-----H--- 385  
Db 4015 SKFRMSFVDSVAVLEMHNLSDVEDSGDYTCENAHNAAGSASSTSLKVKEPPIFRKKPHPIE 4074  
QY 386 -----D-----LO-----Q-----H----- 390  
Db 4075 TLKGADVHLECELOQTPPHVSWYKDKRELRSKGYKIMSENFLTSTIHILNVDAAIGY 4134  
QY 391 -----N-----KEI-Y-TI-----K 398  
Db 4135 QCKATNDVGSDDTCVGSIALKAPPRFVKKLSLDISTVVGKEVQLOTTIEGAEPISVWFKDK 4194  
QY 399 -----W-----S----- 400  
Db 4195 GEIVRESDNIMWISYSENATLQFSRVEPANAGKYTCQIKNDAGMQECFATLSVLEPATIV 4254  
QY 401 -----P-----TG-----P-----G-T-----N-----N-- 408  
Db 4255 EKPSIKVTGTGTCTLECTVAGTPELSTKWKDKGKELTSDNKYKISFFNKKVGLKINVA 4314  
QY 409 -----P-----N-----A-----N-L-----M-----L--- 415  
Db 4315 PSDSGVYSEVQNPVGKDSCTASLQVSDRTVPSPFTRKLEKETNGLSGSSVWMECKYVGP 4374  
QY 416 -----ASAS-F-----DS----- 422  
Db 4375 PISVSWFHEGNEISSGRKYQTTLTNTCALTVNMLESDSGDYTCIATNAGSDECSAPL 4434  
QY 423 -----TVR-----L-----W----- 427  
Db 4435 TVREPPSFVQKPPDMVLTGTNTFTSVYKGTTPPFVSWFKGSSSELVPGDRCNVLSDSV 4494  
QY 428 -----DVD-R-G-ICI-----H----- 436  
Db 4495 ABLELFDVDTSSQGEYTCIVSNEAGKASCTTHLYIKAPAKFVKRLNDYSIEKPKPLILEG 4554  
QY 437 -----TLT-----K-----H-Q-----B-P---V---YS-V--- 448  
Db 4555 TFGTGPISVTWKKNGINVTSPQRNITTEKSPILEIPSSSTVEDAGQYCNVYENASGKD 4614  
QY 449 -----A-----F-----SP-----D-----G-----R---Y-- 456  
Db 4615 SCSAQILILEPPYFVKQLEFPVKVSGDSASLQCLAGTPEIGVSWYKGTCLRPPTTYKM 4674  
QY 457 -----L-----A-S-G-----SF-----D-K- 464  
Db 4675 HFRNNVATLVFNQVINDSGEYICKAENSVEGVSASTFLTVOBQKLPSPFSRQLRVDQET 4734  
QY 465 -----C-V---HI---W-----NT-----QV-C 474  
Db 4735 VGLPVVFDCAISGSEPISVSWYKDGKPLKDSPNVQTSFLDNTATLNI FKTDRSLAQYSC 4794  
QY 475 -----L-----H----- 476  
Db 4795 TATNPIGSASSARLILTEGKNPPFPDRLAPDAVVGESADFECHVTGTQPIKVSWAKD 4854

QY 477 -----YL-N-----GQ-----V-----L----- 483  
Db 4855 SREIRSGRGYQISYLENSAHLTVLKVDKSGQYTCYAVNEVGKOSCTAQLNIKERLIPP 4914  
QY 484 -----L-----N-----L-GR-----S-----I-----C----- 491  
Db 4915 SFTKRLSEVTEETEGNSFKLEGRVAGSQPITVWYKNNIEIOPTSNCEITPKNNTLVQV 4974  
QY 492 -----LYT-----L-----P-----H----- 498  
Db 4975 RKAGMNDAGLYTCKVNDAGSALCTSSIVIKPKPPVFDQHLTPVTVSEGEVQLSCHV 5034  
QY 499 ----- 498  
Db 5035 QGSEPIRIOWLKAGREIKPSDRCSFSGTAVLELRDVAKADSGDYVCKASNVAGSDTT 5094  
QY 499 -----L-----VV-----IP----- 503  
Db 5095 KSKVTIKDKPAVAPATKKAADVGRGLFFVSEPOQSIRVVEKTTATTAIAKVGDDPIPNVWK 5154  
QY 504 -----L-----VA-----L----- 507  
Db 5155 GKWRQLNQGGRVEIHOQDEAKLEIRDTTKTDSGLYRCVAFNEHGBIESNVNLQVDERKK 5214  
QY 508 -----LE-L-----VLK 514  
Db 5215 QEKIEGDLRAMLKKTILK 5233

RESULT 2  
T20992  
hypothetical protein F15G9.4a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T20992; T24733  
R:Sulston, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19355  
A:Accession: T20992  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-5175 <WIL>  
A:Cross-references: UNIPROT:Q810L3; EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a  
A:Experimental source: clone F15G9  
R:Kershaw, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19929  
A:Accession: T24733  
A>Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-5175 <W12>  
A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a  
A:Experimental source: clone T09B9  
C:Genetics:  
A:Gene: CESP:F15G9.4a  
A:Map position: X  
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3; 1252/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2981/3; 3033/1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5077/1

Query Match  
Best Local Similarity 8.8%; Pred. No. 5.5e-53;  
Matches 453; Conservative 49; Mismatches 8; Indels 4660; Gaps 389;  
QY 1 M-----SI-----SS-----D-----E----- 7  
Db 1 MGRSPSWLYGVGLLLATTCSVNDKNDPTGKSLAFVFDITGSMFDDLQVREGAAK 60  
QY 8-----V-----N-----F-----L-----VY----- 13  
Db 61 IFKTVMAQREKLIYINWVPHDPIYLGEIINTDSTTFMRQLSKVYVHGDCPEKLTG 120  
QY 14-----R-----Y-----L-----QE-----SG-----F-----SHSAF-T 27

Db 121 ILKALQISLPSSFIYVFTDARSKDVHLEDEVLTIQEKSSVVFMTGDCGNRTHPGERT 180  
QY 28 -----FG-I-----KS-----H-ISO-----S-NI----- 39  
Db 181 YEKIAAASFGQVHLEKSDSVTVLEYVRHAVKQKVHLMYFARERGGTVSRNIPVDKHL 240  
QY 40 -----N-----GA-----LV-----P-P----- 46  
Db 241 ELTISLSDKDDNDLIVLRDPEGRTVDKRLYSKEGGTIDLNKVKLIRLKPSPGVWTV 300  
QY 47 -----A-----A-----LI-S-I----- 52  
Db 301 NTNSRLKHTIRVFGHGVDFKYGFASRPLDRPLRIELARPRVLNQDTYLLINMTGLIPP 360  
QY 53 -----I-----Q-----KGL-----Q-Y-----VE-----AE----- 63  
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYDEDDTEFMR 420  
QY 64 -----V-----S-I-----N-----ED-----GTLF----- 73  
Db 421 AIGSVIVGPPRAFSPHIEQFVGRDLNLSCTVESASAYTIYWKGTGEDIIGGLFVHND 480  
QY 74 -----D-----G-----R-P-I-----E----- 79  
Db 481 TSVWTIPELSLKDAGEYECRVISNNGYSVKTRVETRESPPFIFGVNRVSVPLGEAFLH 540  
QY 80 -----S-----L-----SL-I-----DA----- 86  
Db 541 CSTRSAGEVEIRWTRYGATVFNPNTERNTNPTNGTLKIHVTRADAGVVECMARNAGG 600  
QY 87 -----VM-----P-DV-----V----- 92  
Db 601 RMRDLIMEPPSVKTPQDVIYFNMREGVNLSCAMGDKPKPEVHWYFKGRHLLNDYKQVG 660  
QY 93 Q-----T-----R-----Q-QAYRD-----KLA-----Q----- 105  
Db 661 QDSKFLYIRDATHDEGTVECRAMSGAQA-RDITDLMLATPPKVEIIQNMVGRGDRV 719  
QY 106 -----Q-----Q-----A-----A-----A-----A----- 111  
Db 720 SPECTIRKPKPKIRFWKNGKDLKPDYIKINEGOLHIMGAKDEADAGAYSCVGENMAG 779  
QY 112 -----A-----A-----A-----A-----A-----A----- 114  
Db 780 KDQVANLSVGRVPTTIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNP 839  
QY 115 -----A-----A-----A-----ASQ----- 120  
Db 840 RYTQLADGNLLITDAQIEDQGOFTCIARNTYQSQSQTTLMTVGLVSPVLGHVPPEQLI 899  
QY 121 -----Q-----GS----- 123  
Db 900 EGQDLTLSCVVVLGTPKPSIVMKDKKPVEEGFTIKIEGSSLLRLRGNGPKDEKYTCI 959  
QY 124 -----A-----K-----N-----GE----- 128  
Db 960 AVSPAGNSTLHINVLKPKPEVYKPEGGIVFKPTISGMDEKHVAVVNSTHVDLGEFA 1019  
QY 129 -----NT-----A-N-G----- 133  
Db 1020 IPCVSGSTPPPIITWYLDGRPTPNRSRDTFTADNTLIVRKADKSYSGVYTCQATNSAGD 1079  
QY 134 -----E----- 136  
Db 1080 NEQKTTIRIMTPMISPGOSSFNWVDDLFTIPCDVYGDPKPVITWLLDDKPTEGVNE 1139  
QY 137 -----G-----AH-----TI-A-N-----N----- 144  
Db 1140 DGSLLTIPNVNEAHRGFTTCHAOAAGNDTRTTLTVHTTPTINAENQEKIALQNDIVLE 1199  
QY 145 -----HT-----D-----MME-----V-----D----- 152

Db 1200 CPAKALPPVRLWTEGEKIDSQLPHITIREDGALVLQNVKLENTGVFCVQVSNLAGEDS 1259  
QY 153 -----G-DV-----EIP-----P-----NK-AV----- 163  
Db 1260 LSYLTIVHEKPKLIISVPGVVDVVKFTIIPCRATGVPEVITWNGIDLKMDEKKS 1319  
QY 164 V-----LR-----G-----H-----E-----S----- 170  
Db 1320 VDNLGLRIYADKNDIGNVNCVVTNEAGTSQMTTHVDVQEPPIILPSTQNTNNAVVGDR 1379  
QY 171 -----E-----V-F-----I-C----- 175  
Db 1380 VELKCYEASPPASVTWFRGIAIGDTKGYVVESDGLTIVIOASVEDATIYTCASNEA 1439  
QY 176 -----A----- 176  
Db 1440 GRAEANLQVTVIASPDIKOPDVVTQESIKESHPSLYCPVSNPLPQISWYLNKPLIDD 1499  
QY 177 W-----N----- 178  
Db 1500 KTSWKTSDDKRLHVFKAKITDSGVYKCVARNAAGSKSFQVEVIVPLNLDSEYKKKV 1559  
QY 179 -----PVS----- 181  
Db 1560 FAKGSEVTLGCPVSGFPVQINWVVDGTVVEPGKYGATLSNDGLTLHFDVSVKQEG 1619  
QY 182 -----D-L-LA-----S 186  
Db 1620 NYHCVAQSGNILDIDVELSVLAVPIVGEDDNLVFLGKDISLSCDLQTESDDKTFVWS 1679  
QY 187 --GS-----DS-----T-----A-R----- 194  
Db 1680 INGSBDRPDNVQIPSDGHRLYITDAKPNNGKYMCRVTNSAGKARTLTLVLEPPVVF 1739  
QY 195 -----IW-----N-----LS----- 199  
Db 1740 EPVFEANQKLIIGNPFILOQVGTGNPKPTVIWKIDGNDVDKSWLPDESLSLRIEKLTKG 1799  
QY 200 -----EN-----S--T-----S-GS--T- 207  
Db 1800 SAQISCTAENKAGTASRDPFIQNIAPTFKNEGQDQETIFRESEITILDCPVSUGDFOITW 1859  
QY 208 --Q-----L-VL--R-H--C-----IR 216  
Db 1860 MKQGLPLTENDALFTLDNTRLTLNARHEDIYTCVANNTAGQVSKDFDVVVQVLPKIK 1919  
QY 217 -----EG-----G-----Q-DVP----- 223  
Db 1920 NAVVTLEINEGEBIILTCDAGNPTPTAKWDFNQGLDKPEAVFVNNNHTVWVNNVTKYHT 1979  
QY 224 -----SNK-----DV-----T-S-L--D----- 232  
Db 1980 GYVKVATNKVQAVTINHVTKPRFSGLTESBLTNLRSITLECDVDDAIGVGIS 2039  
QY 233 W-N-----SE-G-TL--L--A-T--GSY----- 245  
Db 2040 WTVNGKPLAETDGVQTLAGRELHIVSAKTDDHGSYACTVNEAGVATKTNLFVQVPP 2099  
QY 246 -----DG----- 248  
Db 2100 TIVNEGGEYTVIENNSLVLPCVETGKPNPVVVTWKDGRPVGLKSVQLSEGOQFKIVHA 2159  
QY 249 -----A-----R----- 250  
Db 2160 EIAHKSICYCMKNDVGTABISFDVDIITRPMIOKGIKNIVTAIKGALPFCFKPIDDDKN 2219  
QY 251 -----IW-----T-----K-D-G-N----- 257  
Db 2220 FKQIILWLNYPIDLEADARITRLSNDRLTLNVNTEDEQYSCRKNDAGENSFDF 2279  
QY 258 -----L-----A--S--TL-----GO--H----- 265  
Db 2280 KATVLVPPTIIMLDKDKNKTAVEHSTVTLSCPATGKPEPDIITWFKDGEALHIENIADIIP 2339  
QY 266 ----- 265  
Db 2340 NGELNGNQLKTRIKBGDAGKYTCESADNAGSVEQDVNVNVIITPKIEKDGIPSDYESQQ 2399  
QY 266 -----K-G-P-----I-- 269  
Db 2400 NERVVISCPVYARPPAKITWLKAGKPLQSDKFVKTSSANGQKLYFLKRLRETDSSKYTCIAT 2459  
QY 270 -----F-----A-----L--K-----W----- 274  
Db 2460 NEAGTDKDRPKVSMVLVAPSFDEPNIVRRITVNSGNPSTLHCPAKGSPSTIITWLKGNAI 2519  
QY 275 --N-----K----- 276  
Db 2520 EPNDRYVFPDAGRQLQISKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPVIDGERREA 2579  
QY 277 -----K-G--N-----FI-----LS----- 283  
Db 2580 VAVIEGSELSFCDNSNSTGVDEWQKGLTIINQDTRLGDSFIQIPSSGKKMSFLSARKSD 2639  
QY 284 -----AG-----V-----DK--T--TI-----I----- 292  
Db 2640 SRYTCIVRNPAEARKLDFFAVNDPPSISDELSSANIQTIVPYYPVEINCVSQSGSPHK 2699  
QY 293 --W-----D--A-----H-----T-----G----- 298  
Db 2700 VYWLFPDDKPLEPDSAAVELTNNGETLKIIVRSQVEHAGTYTCBAQNNVGRKARDFLVRVTA 2759  
QY 299 -----E-----AK-----Q----- 302  
Db 2760 PPHFEKEREBVAVRGDTMLLTCAABESSVPLSSVYVHHADESQNGVITSKYAANEKTLN 2819  
QY 303 --Q-----F----- 304  
Db 2820 VTNIQLDDEGYCTAVNEAGITKFFKLIIVETPYFLDQQLYPIILKRLTLDCSATG 2879  
QY 305 --P--F----- 306  
Db 2880 TPTPILFMKDKRLNESDEVIIIGSTLVIDNPQKEVGRYTCIAENKAGRSEKMMVEV 2939  
QY 307 -----H-----S----- 308  
Db 2940 LLPKLSKEWINVEVQAGDPLTLECFIEDTSGVHITWSRQFGDKGQDLMRAQSSDKSL 2999  
QY 309 ---A-P--A--L-----D-----V-D----- 315  
Db 3000 YIMQATPEDADSYCIAVNDAGAEAVFOVTNTPPKIFGDSFSTTEIVADTTLEIPCT 3059  
QY 316 -----W-----Q----- 317  
Db 3060 EGIPPEISFLLDKPILMPGVYKQDLSLRIDNIKPNQBGRYTCVAENKAGRAEQDT 3119  
QY 318 -----S-----N-----N-----T-----P----- 322  
Db 3120 YVEISEPPRVWMASEVMRVVEGRQTTIRCEVFGNPEPVVWNLKDGEPYTSDDLQFSTKLS 3179  
QY 323 -----A--SC-----S--TD-----M--C-- 330  
Db 3180 YLHLRETLADGGTYTCIATNKAGESQTTDVEVLVPPRIEDEERVLOGKEGNTYVWHCQ 3239  
QY 331 -----I-----HV-----C-----K 335  
Db 3240 VTGRPVVYTWKNGKEIEQFNPVLHIRNATRADEGKYSCIASNEAGTAVADFLIDVFTK 3299  
QY 336 -----L--GO-----D-----RP-----I-K- 343  
Db 3300 PTFETHETTFNIVEGESAKIECKIDGHKPTISWLKGRPFNNMDNIIILSPRGDTLMILKA 3359  
QY 344 -----T-----F-----Q----- 346  
Db 3360 QRFDGGLYTCVATNSYSDSEQDFKNVYTKPYIDETIDQTPKAVAGGEIILKCPVLGNPT 3419

Qy	347	----	G	----	HT	----	N	----	E	----	VN	----	A	I	----	355
Db	3420	PTVTW	KRGDDA	VPNDS	RHTI	VNNY	DLKINS	VTTE	DAGQ	YSCIA	VNEAG	NLTTH	YAAE	VIG		3479
Qy	356	----	K	----		----	W	----		----	D	----	P	----	T	360
Db	3480	KPTF	VRKGN	LYE	VIEND	TTIM	DGVT	SR	PLP	SIS	WFR	GR	KPV	LYD	RSIS	3539
Qy	361	----	G	----	N	----	L	----		----	LA	----	S	----	C	368
Db	3540	NKAK	LSDGG	KV	CRAS	NEAG	TS	DIDL	ILK	ILV	PPK	DKS	NI	GN	PLA	3599
Qy	369	----	D	----	M	----		----		----		----		----		371
Db	3600	GIPQ	PVIT	WTK	NGM	IND	TSR	VILA	QON	TE	FG	IE	NV	QV	TD	3659
Qy	372	----		----	TL	----	K	----	I	----	W	----	S	----	M	378
Db	3660	SLDV	SPPE	FI	HGT	OPT	K	REG	DT	IT	LC	PIA	ED	IA	DQ	3719
Qy	379	----	K	----	Q	----		----		----	D	----		----		381
Db	3720	NVDI	SDGR	KLT	TS	QAS	LEN	AG	LY	TC	IA	LN	R	GE	AS	3779
Qy	382	--N	----	C	V	----	H	----	D	----	L	----	Q	----	H	390
Db	3780	AVNO	PTIM	CA	VT	CH	PP	PS	IK	WL	KNG	KE	V	TD	DE	3839
Qy	391	--N	----	KE	----		I	----	Y	----	TI	----		----	KW	399
Db	3840	AEND	AGV	KE	LE	MV	D	V	TP	PV	SV	K	SN	P	I	3899
Qy	400	--SP	----	T	----	G	----	P	----	GT	----			----	N	407
Db	3900	FDS	PD	G	ARI	SL	K	G	AR	D	I	P	HL	K	T	3959
Qy	408	----		----	N	----	P	----		----		----		----		409
Db	3960	MS	PL	PA	Q	S	LT	L	Q	LA	Q	K	P	V	Q	4019
Qy	410	----	N	----	A	----	NLM	----	LA	----		----	S	----		417
Db	4020	G	V	T	C	Y	A	E	N	A	G	S	D	L	M	4079
Qy	418	----		----	A	----	S	----		----		----		----		421
Db	4080	NG	R	V	T</											

[illegible]

### RESULT 3

T43290

hemicentin precursor - *Caenorhabditis elegans*  
143250

C:Species: *Caenorhabditis elegans*

C/species: *Caenorhabditis elegans*  
C/Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text change 09-Jul-2004

C; Date: 11-Jul-2000 #sequence 1618  
C: Accession: T43290: T20993: T24734

R:Vogel. B.E.: Hedgecock. E.M.

K; Vogel, B.E.; Hedgescock, E.M.  
submitted to the EMBL Data Library. June 1998

**A-Description:** Hemocentin is required for hemostasis in Drosophila. The gene encoding hemocentin was identified by positional cloning. The protein consists of 68 amino acids and is highly conserved among insects. It is secreted from the prohemocytes of the fat body and acts as a humoral factor to induce the differentiation of hemocytes. The gene has been submitted to the EMBL Data Library, June 1990.

A;Description: Hemiscantin  
A:Reference number: 722396

A; Reference number:  
A: Accession: T42290

A;Accession: T43290

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-5198 <VOG>

A; Cross-refer

R; Sulston, J.

submitted to the EMBL Data

A;Reference number:

A;Accession: T20993

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-5198 <WIL>

**A; Cross-references:** EMBL: Z47068; P

## A; Experimenta

submitted to the EMBL Data Library, December 1994

A:Reference number: Z19929

A:Accession: T24734

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5198 <W12>

A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CBSP:F15G9.4b

A:Experimental source: clone T09B9

C:Genetics:

A:Gene: him-4; F15G9.4b

A:Map position: X

A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;

2512/2; 2699/3; 2759/1; 2852/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1

1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/

Query Match 70.9%; Score 2586; DB 2; Length 5198;

Best Local Similarity 9.9%; Pred. No. 6.2e-53;

Matches 423; Conservative 70; Mismatches 19; Indels 3770; Gaps 370;

QY 1 M-----SI-----SS-----D-----E----- 7

Db 1 MGRSPSLYGLVGLLLATTCSSVNDKNDPTCKSSLAFFVDITGSMFDDLVQVREGAAK 60

QY 8 ----V-----N-----F--L--VY----- 13

Db 61 IFKTVMAQREKLIYIMVPHDPYLGEIINTDSTYFMRQLSKVYVHGGGDCPEKTLTG 120

QY 14 -----R--Y-----L--QE--SG--F-----SHSAF--T 27

Db 121 ILKALQISLPSSFIYFTDARSQYHLEDEVLNTIOEKQSSVVFVMTGCGNRTHPGFRT 180

QY 28 -----FG--I-----KS-----H--ISO-----S--NI----- 39

Db 181 YEKIAAASFGVPHLEKSDVSTVLEVRHAKVKVHLMYAEARGGTVSRNIPVDXKLS 240

QY 40 -----N-----N-----GA-----LV-----P--P----- 46

Db 241 ELTISLGDKSDNLDIVLRPEGRTVDKRLYSKEGGTIDLKNVKLRLKDPSPGVWTV 300

QY 47 -----A-----A-----A-----LI--S--I----- 52

Db 301 NTNSRLKHTIRVFGHGAVDKFGFASRLDRLELARPVLNQDTYLLINMTGLIPPGTV 360

QY 53 --I-----Q-----KGL-----Q--Y-----VE-----AB-- 63

Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYDEDNFEMRIAPT 420

QY 64 -----V-----S--I-----N-----ED-----GTLF----- 73

Db 421 AIGSVIVGGPRAFPSPFHQEFVGRDLNLSCTVESASAYTIYWKGTGEDIIGPLFYHNTD 480

QY 74 -----D-----D-----G-----R--P--I-----E----- 79

Db 481 TSVWITIPELSKDAGEYECRVISNNNGYVSKTRVETRESPEIFGVRNVSVPLGEAAFLH 540

QY 80 -----S-----L--I-----DA----- 86

Db 541 CSTRSAGEVEIRWTRYGATVFGNPNTNPTNGTLKIHVTRADAGVYECMARNAGGMT 600

QY 87 -----VM-----P--DV-----V----- 92

Db 601 RKMRLDIMEPPSVKVTFDQVYFMREGVNLSCAMGDKPKPEVHWYPKGRHLLNDYKYQVG 660

QY 93 Q-----T--R--Q--QAYRD-----KLA-----Q----- 105

Db 661 QDSKELYIRDATHDEGTGYECRAMSAGQA--RDTTDLMLATPPKVBIIQNMVMVGRGDRV 719

QY 106 -----Q-----Q-----Q-----A-----AAA-----AA-- 113

Db 720 SPECKTIRKPHKPKIRWFKNGKDLIKPDDYIKINEGQLHMGAKEDAGAYSCVGENMAG 779

QY 114 --AAAA-----AS-----Q-----Q-----GS--A--KN-- 126

Db 780 KDVQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPPPEIEWQKGNVLLATLNNP 839

QY 127 -----GEN-----T--A--N----- 132

Db 840 RYTQIADG--NLLITDAIEDQOQFTCIARNTYQQSQSTTLMVTGLVSPVLGHVPPEQL 898

QY 133 -----G-----EE-----N--G----- 137

Db 899 IEGQDLTSLCVVVLGTPKPSIVMIKDDKPVESGPTIKIEGGGSLRLRGNPKDEGKYTC 958

QY 138 -----A-----H-----TI-----A--N--NH----- 145

Db 959 IAVSPAGNSTLHNVQLIKKPFVYKPEGGIKFTPTISGMDEKHVAVVNSTHDLVDEGCF 1018

QY 146 -----T-----T----- 146

Db 1019 AIPCVCVSGTPPPIITWYLDGRPIITPNSRDTVTADNTLIVRKADKSYSGVYTCQATNSAG 1078

QY 147 D-----M--M-----E-----VD-----GD-----V-- 155

Db 1079 DNEQKTTIRIMTWPMSIPGQSSFNWVDDLFTIPCDVYGDPKPVIITWLLDDDKPFTEGVVN 1138

QY 156 E-----IP-----P-----N--K--A-----VVL 165

Db 1139 EDGSLTIPNVNEAHRGTFTCHAQNAAGNDRTVTLTHTTPTINAENQEKIALQNDIVL 1198

QY 166 -----R-----R-----G-- 167

Db 1199 ECPAKALPPVRLWTYEGEKIDSLIPHTIRDEGALVQNVKLENTGVFCVQVSNLAGED 1258

QY 168 -----HE-----SEV-----F--I--C--A-----WNP--V----- 180

Db 1259 SLSYTLTVHEKPKILSEVPGVVDVVGKFTIEIPCRATGVPEVIRTNKNGIDLKMDEKFF 1318

QY 181 S--D--L--L--AS--G-----SG-----D-----S-----TA----- 193

Db 1319 SVDNLGTLRIYEADKNDIGNYCVVTNEAGTSQMTTHVDVQBPPIILPSTQTNNTAVVGD 1378

QY 194 RI-----W-----NL-----S--ENST-----S-- 204

Db 1379 RVELKCYEASPPASVTFWFRGIAIGTDTKGYVVEDGLTVIQSASVEDATIYCKASNP 1438

QY 205 -G-----S-----TO-----L--V-----L-- 211

Db 1439 ACKAEANLQVTVIASPDIKDPDVVTQESIKESHPSLYCFVFSNPLPQISWYLNDRPLID 1498

QY 212 -----R--H-----CI--R----- 216

Db 1499 DKTSWKTSDDKRLHVFRAKITDSGVYKCVARNAAGEGSKSFQVEVIVPLNLDSEYKXK 1558

QY 217 ---EG-----G--Q-----D--V--P-----SN----- 225

Db 1559 VFAKEGEVTLGCPVSGPPVQINWVVDGTVVVEPGKKGATLNSDGLTLHFDSVSVKQE 1618

QY 226 -----A-----TG-----SY--D----- 246

Db 1619 GNYHCAQSKGNLIDIVELSVLAVPIVGEDDNLEVLFLGKDISLSCDLQTESDDKTFVW 1678

QY 234 --N--SE-----G-----G-----TL--L----- 240

Db 1679 SINGESDRPDNVQIPSGHRLYITDAKPNNGKYMCRVTNSAGRAERTLTLDVLEPPVF 1738

QY 241 -----A-----A----- 246

Db 1739 VEPVFEANQKLGNNPIILQCVQTNPKPTVIWKIDGNDVDSWFLDESLLRIEKLGT 1798

QY 247 -----G-----F--A-----R-----I-- 251

Db 1799 KSAQISCTAENKAGTASRDPFFIQTAAATFKNEGDEQETIFRESEITLDCPVSLGDFQIT 1858

QY 252 W-----T-----T-----KO----- 255

Db 1859 WMKQGLPLTENDAIFTLDNTRLTILNANRDHEDIVTCVANNTAGQVSKDFVWVQVLPKI 1918

QY 256 -----GN-----L--A-----S--- 260  
Db 1919 KNAVTLLEINEGBEIILTCAENPTPTAKWDFNQGLDPKEAVFVNNHTVVVNNVTKYH 1978  
QY 261 T-----L-----GO-----H--K----- 266  
Db 1979 TGYYKCATNKVGOAVKTVNHVTKPRFESGLTESELTVNLFRSITLCECDVDAIGVI 2038  
QY 267 -----G-P-----I-----F----- 270  
Db 2039 SWTVNGKPFLEAETDGVQTLAGRFLHIVSAKTDHGSYACTVNEAGVATKTNLFVQVP 2098  
QY 271 -----AL-----K-----W----- 274  
Db 2099 PTIVNEGGEVTVIENNSLVLPCEVTGKPNPVVTWKDGRPVGDLKSVQVLSQOQFKIVH 2158  
QY 275 -----N-----K-----KG----- 278  
Db 2159 AETAHKGSYICMAKNVGTAEISFDVDIIITRPMIQKIKNIVTAIKGALPFPKPIDDDK 2218  
QY 279 NF-----I-----LS-----AG----- 285  
Db 2219 NFRGQIWLRYOPIDLEADARITRLSNDRLTLNVNTENDRGQYSCRKNDAGENSFD 2278  
QY 286 -----V-----D-KT-----TI-----I-W-D-A-H----- 296  
Db 2279 FKATVLVPPTIIMLDKDKNTAVEHSTVLSCPATKPKPEPDITWFKDGEAHIENIADII 2338  
QY 297 -----T-----G-BA-----K-----Q 302  
Db 2339 PNGELNGNLKIRIKEGDAGKYTCRADNSAGSVEQDVNVNVTIPKIEKDGPSPYESQ 2398  
QY 303 Q----- 303  
Db 2399 QNERVWSCPVYARPPAKITWLKAGPLQSDKEVKTSAHQKLYLFKLRDTSKYTCIA 2458  
QY 304 -----F-----P-F-----H----- 307  
Db 2459 TNEAGTKRDFKVSMLVAPDFPNVIRRVITVNSGNPSTLHCPAKGSPSTITWLKDGNA 2518  
QY 308 -----S-----A-----P----- 310  
Db 2519 IEPNDRVFPFDAGRLQISKTEGSDQGRYTCIATNSVSGDLENTLEVIIPPIDGERRE 2578  
QY 311 --A-----L-D-----VD-WQ-----S-N-NT-----F-A----- 323  
Db 2579 AVAVIEGFSSELPFCDNSNSTGVDEWQKGLTINQDTRLGDSFTIQIPSSGKKMSFLSARKS 2638  
QY 324 -S-----C-----S-TD-W-----CI-----H- 332  
Db 2639 DSGRYTCIVRNPAEARKLFDFAVNDPPSISDELSSANTQIVPVYVPEINCUSVSGSPHP 2698  
QY 333 -----V-----CK-----LG-Q-D----- 339  
Db 2699 KVVYLFDDKPLEPDSAAELTNNGETLKI VRSQVEHAGTVTCEAONNVGKARKDFLVRVT 2758  
QY 340 -----R-----P-----I-----KT- 344  
Db 2759 APPHFEKEREVVARVGDWMLTCNAESSVPLSVVYHHADESQVNGVITSKYAANEKTL 2818  
QY 345 -----F-Q----- 346  
Db 2819 NVTNIQLDDEGFFYCTAVNEAGITKFFKLVIVETPFYLDQKLYPIILGKRLTLDCSAT 2878  
QY 347 G-----H-----T-N-EV-----N-A----- 354  
Db 2879 GTPPTILFMKGDKRLNESDEVDIIGSTLVIDNPKEVEGRYTCIAENKAGRSEKDMVVE 2938  
QY 355 I-----K-W-----DP-----T-----G-----N 362  
Db 2939 VLLPPKLSKEWINVEVQAGDPLTLECFIEDTSGVHITWRSQFGKQOLDMRAQSSDSKSK 2998

QY 363 LL-----A-SC-----S-----D-----DMTLKI--- 375  
Db 2999 LYIMQATPEDADSYSCIAVNDAGBAEAVFQVTVTPKIFGDSFSFSTEIVADTTLEIPCR 3058  
QY 376 -----W-----SM-----KQ-----DN-----CV----- 384  
Db 3059 TEGIPPPISWFLDKGPILFEMPQVTVYKQGLDLSLRIDNIKPNOBGRYTCVAENKAGRAEQD 3118  
QY 385 ----- 384  
Db 3119 TYVEISEPRVVMASEVMRVRVEGRQTTIRCEVFGNPEPVVNMWKDGEPTYDLSLQFSTKL 3178  
QY 385 --H-----D-L-----O-----Q-----H- 390  
Db 3179 SYLHLRETTADGGTTCIATNKGESQTTTDEVLVPPRIEDEBRLVQKSGNTYMHVC 3238  
QY 391 -----N-KEI-----Y----- 395  
Db 3239 QVTGRPVVYVWKRNKGKEIEQFNPVLHIRNATRADEGKYSCTIASNEAGTAVADFLIDVFT 3298  
QY 396 -----T-----I-K-W-----SP----- 401  
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRPFNMNDIILSPRGDTILMLK 3358  
QY 402 ----- 403  
Db 3359 AQRFDGGLYTCVATNSYGDSEQDFKVVVTKPYIDETIDQTPKAVAGGEIILKCPVLGNP 3418  
QY 404 -P-----G-----T--NN----- 408  
Db 3419 TPTVTWKRGDGDAVPNDSRHTIVNNYDLKINSVTTEDAGOVSCIADVNEAGNLTHYAAEVI 3478  
QY 409 --P-----N-----AN-L-M-----LASAS-F----- 420  
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLFSISWFRGDKFVYLYDRYSISPDGSHIT 3538  
QY 421 -----D-----S--T-----V-R--L----- 426  
Db 3539 INKAKLSGDKYICRASNEAGTSDDLILKILVPPKIDKSNIIGNPLAIVARTIYLECPI 3598  
QY 427 -----W-----DV--D-R-----GI-----CI-----H- 436  
Db 3599 SGIPQPDVITWTKNGBDINMTDSRVILAONNETFGIENVQVTDQGRYTCATNRGKASHD 3658  
QY 437 ---TLT--K--H-Q-----E-----P----- 444  
Db 3659 FSLDVLSPPEFDIHGTQPTIKREGDTITLTCPKLAEDIADQVMDVSVWTKDSRALDGLT 3718  
QY 445 --V-----YS-VA-----F-----SP-----D----- 453  
Db 3719 DNVDISDDGKRLTISOASLENAGLYTCIALNRAGEASLEFKVEILLSPVVIDISRNDVQPO 3778  
QY 454 -----G-----R-----Y-L-AS--G-S- 461  
Db 3779 VAVNQPTIMRCVATGHPPPSIKMLKNGKEVTDENIRIVEQGOVLQILRTSDSHAGKWSC 3838  
QY 462 -----F-----D--K-----C----- 465  
Db 3839 VAENDAGVKELEMLVDVFTPPVSVKSDNPIKALGETITLFCNASGNPYPQLKWAAGGSL 3898  
QY 466 -----V-H-----I----- 468  
Db 3899 IFDSPDGARISLKGARLDIPLHKKTDVGYTCQALNAACTSEASVSDVLVPEINRDMI 3958  
QY 469 -----W-N-----T-Q----- 472  
Db 3959 DMSPLPAQOQSLTLOCLAQKFPVQMRWTNGTALTHTSTPGITVASDSTFIQINNVSLS 4018  
QY 473 --V--C-----LHY-----L-NG--QV-----L- 484  
Db 4019 KGYYTCYAENVAGSDNLMYNDVQVAPVISNGTQVIEGELAVTECLVEGYPAPQVSWL 4078  
QY 485 --N-----L--GR-----S-I-C-----LYTL----- 495



Db 4079 RGNRVETGVQGVRYVTDGRMLTIEARSLSIDSYLCSATNEAGSAQAYTLEVLVSPKI 4138  
QY 496 -----P-----H-----497  
Db 4139 IYSTPGVLTTPSGSKFSLCAVGVGPDPIISWTLNGNDIKDENGHTIGADGTLHIEKAE 4198  
QY 498 --HLV-----V-I-P-----LV-----AL--IE-----509  
Db 4199 ERHLIYEKADNAGADTLEFPQTVAPKISTSGNRYINGSEGTETVIKCEIESSESEF 4258  
QY 510 -----L-----LV-----LK 514  
Db 4259 SWSKNGVPLPSNNLIFSEDIK 4280

RESULT 4  
T30192  
probable peptide synthetase - Aureobasidium pullulans  
C:Species: Aureobasidium pullulans  
C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 09-Jul-2004  
C:Accession: T30192  
R:Peery, R.B.; Thorneswell, S.J.; Tobin, M.B.; Skatrud, P.L.  
submitted to the EMBL Data Library, January 1997  
A:Description: Discovery of an MDR-like gene adjacent to a peptide synthetase in Aureoba  
A:Reference number: Z20767  
A:Accession: T30192  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-10797 <P>E>  
A:Cross-references: UNIPROT:O94116; EMBL:U85909; NID:g4099310; PID:g4099313; PIDN:AAD005  
C:Genetics:  
A:Introns: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2  
C:Keywords: carrier protein  
F:1618-1688/Domain: acyl carrier protein homology <ACP1>  
F:3682-3752/Domain: acyl carrier protein homology <ACP2>  
F:5615-5685/Domain: acyl carrier protein homology <ACP3>  
F:7503-7573/Domain: acyl carrier protein homology <ACP4>  
F:9683-9752/Domain: acyl carrier protein homology <ACP5>

Query Match 70.8%; Score 2585; DB 2; Length 10797;  
Best Local Similarity 8.8%; Pred. No. 3,1e-52;  
Matches 457; Conservative 42; Mismatches 10; Indels 4680; Gaps 387;

QY 1 M-----SI-S-----S-----D-----6  
Db 4302 MLNTQVIOFGSGIGKSGILYMGVGTAAIHCTLPQGFVDLPAGTIGIPDITVSCFIV 4361  
QY 7 -----E-V-----N-----FLV-----YR-----14  
Db 4362 KPTESTKASQLEILPIGELVIGGHQADGYNLREBQTRAF-VTHPKFGGLYRTGD 4420  
QY 15 -----Y-----L-Q-----E-----SG-----F-----21  
Db 4421 KARLHNGTLECYGRISGQVKGQRVELGEIHAASKAGGCHAVIASVIGLLVFCI 4480  
QY 22 --SH-----SA-----F-----T-F-----28  
Db 4481 GDPHRVSSKDIKSAQKWPAYMIPSDIVLDDFPYLPSPGVKDKRLETDYNSNTAQHVS 4540  
QY 29 G-----IKS-----H-----I--SQ-----36  
Db 4541 GSSDLSENAREITRIESVLGVIDSHSTLSAAGLDSLRAIQVASQLRQGCADIGALEL 4600  
QY 37 -----SNIN-----G-----41  
Db 4601 LSVSNVLADELVRKADESNINDSEKWKQTVHELRSVERDEPKAFVSGIEDVLPC 4660  
QY 42 -----A-LV-----P-P-----A-----AL-----IS-51  
Db 4661 TPLQDAMLVETAKRPOAYCNELELTVSPKIPVERVRQALFALAQORHTALRSGFMPSGVSH 4720  
QY 52 -----II-----Q-K-G-----L-QY-----VEAE--V-----64

Db 4721 CAYTQVIMKTLVTSQFAHVKSFTTGWSTNRETLLRPLHFQYKCSGAEAILVAIHHALY 4780  
QY 65 ---S---I-----NE-----D--G-70  
Db 4781 DOWSVEVILEDLTLQERTPTSPFGAVNKPENLRSEDOTSHLDWFGEYLSDTVPGR 4840  
QY 71 -----TL-----F-----D-74  
Db 4841 LPNLSFKMPPQPOLOSIQHTIEMDMETLQAAHSYCSAHVFFQAAAYAILLGYMGTEDT 4900  
QY 75 -----GR-P--IES-----LS-L--I-----DA-----VM-----88  
Db 4901 VFGTVFSGRTLPIVEIESMVGPLSLTPRINTLESRKFSVLSRLQEDNKRIMRHSMTS 4960  
QY 89 -----P-----D-V-V-Q-T-----R-----Q-97  
Db 4961 LADIKKACGFNPGEAVFDSIFVMQETARPDARAQTLNLNVEAHYLEFNLTLLEPTQOG 5020  
QY 98 ---A-Y-----R-----DKL-----A--Q-105  
Db 5021 VTKATYQSSLLPLQHVKTLLQQLDALVKIVVAPETHNNEISDQLPISVLSVANSEPOS 5080  
QY 106 -----Q-----Q-107  
Db 5081 FVYKAGLSLVENHALNNSGGLALVFAHDIREGTSRMESLTYGELNTRANQLANYLISOG 5140  
QY 108 A-----AA-----AA-----A-----113  
Db 5141 AKRDELICVMEKSVSLYLILAAVAVKAGCYLPLVPETPAARIRQILAEADVFKCLTDSS 5200  
QY 114 A--A--A-----A-----A-A-----118  
Db 5201 MAPVIADVSRCHIMNVDTDCSAQSGTGPQLDFKPTDIAYAVFTSGTTCKPKGVLTQEN 5260  
QY 119 -----S-Q--Q-----Q-----G--SA-----K-125  
Db 5261 ILSNLEVLKSIYPPEGSRLQACNQAFDVSVEIFFTWYTGWCLCSASKVMFRDFEKA 5320  
QY 126 N-----GE-----128  
Db 5321 INELEITHLSLPTVAALTDPAPHRVKFLVTAGEAVTHHVHAGMAGKGLYQYGPSETT 5380  
QY 129 N--T--A-----N-GE--EN-----GA-----138  
Db 5381 NICTNSAVESDHVINNIQAPFENTSAFVLTQGDFFQLVPLGGLGELCFGGQVFRGYQN 5440  
QY 139 -----H-----TI-----141  
Db 5441 MPELTSKIINHPNYGRIYRSGDLGRLLPDGTILIQGRDQDKIRGORIELGEISGCLL 5500  
QY 142 ---AN-----N-----H-----145  
Db 5501 QPPSVQNCIAIEVIKTADKERLMAFIPSGYSKDSYSILOPKNLEELIKSIYAHADNLP 5560  
QY 146 -----T--DM-----M--E--150  
Db 5561 AYMWPDALVPVSAIPQTSQGIKDKRLASDGSALTVEDLNAYSRGADDETSELSTEQO 5620  
QY 151 -----VD-----G-D-V-----E-----I--P-----158  
Db 5621 LASALADTLQMSQTSIGRSTSPFALGLDSVSAIRLATNLKKEYGVSIDVSYILKRPTIAR 5680  
QY 159 --P-----NK-----AV-----VL-----165  
Db 5681 LAPLLGGESSKOTNEPVTADCEAAVGSYLHDSVSVQLHEHGQTVSQVLPCTPLQEAMLS 5740  
QY 166 R--G-----H-----E-----S-E--V-FI-----174  
Db 5741 ARDTSGSSAYRNKTLFSLHGSVDKLCACWEMQLQRHDLRTIFLSTEDSRFPVQAVLSQ 5800  
QY 175 -----C-----A-----W-----177

Db 5801 WTLPMQECDDIPQLSTLLDSAKAGGDSIVDHSPPWKIOVYRSESTVYLLDMHLYDA 5860  
QY 178 ---N---PVS---D 182  
Db 5861 NAMSLLYEVEQLYKQSLASAPVSKFPLNFMISTSVBEADALFRDQLREFVKKPFRD 5920  
QY 183 ---L---LA---SG- 187  
Db 5921 VKSGFGTITGRNLNYSKPMVETFLSKHSTTMLSITQAMWKTLLAASQSYSDVCCGNVSGR 5980  
QY 188 --- 187  
Db 5981 SVPVDGIESLVAPCFNTIPVRVDSKHSRNLGLVKALQVRNIDSLPYQLTPLRRIOAQAG 6040  
QY 188 ---SG-D---ST--- 192  
Db 6041 TNGKRLFDSLVLLOQDPTDLDLSAIMELEGESGVMYMTSGTGLPKAVCLSHRAVTQSLL 6100  
QY 193 A---R---I-W---N--- 197  
Db 6101 AHDRTFSPSRFLQFASPTDFDSVFEIFFPWYRGATLVSVERNLLGDLPGTITSLNIDA 6160  
QY 198 ---L---S---EN---S---T--- 205  
Db 6161 AELTPSVAASLVHRHENVPTLRALLTIGEMLNTQVIOQFGSGIGKSGILYGMVGPTEAAI 6220  
QY 206 ---ST--- 209  
Db 6221 HCTLPQFGVDLPAGTIGIPDPTVSCFIVKPTSTKHSQLEILPITGEIGELVIGHOLA 6280  
QY 210 ---V---L---R-H---C-I---R--- 216  
Db 6281 DGVLNREEQTRAAFTVHPKFGGLYRTGDKARLHRNGTLECYGRISGQVKLRQORVELGE 6340  
QY 217 E---GG---Q---D-V--- 222  
Db 6341 IEHAASKAGGCHAVIASVIGLLVLCFIDGPHRVSSKDIKSAQKWLPAWMPISDIVLDD 6400  
QY 223 ---P---SN---KD---V---T--- 229  
Db 6401 DFPYPSGKVDKRLFTDYNSTAOHVSGSSDLSENAREITRIESVLGVSIDHSTLSA 6460  
QY 230 ---SL---D---W--- 233  
Db 6461 AGDLSLRATQASQLRQGCADLGALELLSVSNVLALDELVLRAKADENINDNDSEKKWQ 6520  
QY 234 --- 235  
Db 6521 TVHELRSVVERDPESKAFVSGIBEDVLPCTPLQDAMLVETAKRPAQYCNELRLTVSPKIPV 6580  
QY 236 E---G---TL---LA---TG--- 243  
Db 6581 ERVRQALFALAQRHTALRGFSPGVSHCAYTQVIWKTLTVSQFAHVKSFTTGWSVTNRE 6640  
QY 244 ---S---YD---G--- 247  
Db 6641 TLLRPLHFQYKCSGAELVAITHALYDOWSVEVILEDTLLQNERTPERSFGAVNK 6700  
QY 248 -FA-R---I-W---TKD---G---NL--- 258  
Db 6701 FFNLRSEDOTSHLDFWGEYLS-DVTPGRLPNLSPKMPPQPLQSIQHTIEMDMETLRQA 6759  
QY 259 ---A---S---TL---G--- 263  
Db 6760 AHSYSCSAHVFOAAVAILLGFYMGTEDTVFGTSGRTLPIVEIESMVGPLSLTLPTRI 6819  
QY 264 ---Q---H---K-G-P---IFALKW--- 274  
Db 6820 NTLESKFSFVLSRLQEDNKKIMEHSMWLSLADIKKACGNFGEAVFDSIFV-WQETARP 6877  
QY 275 ---N---K 276  
Db 6878 DARAQTLNLNVEAHYLEFNLTLEBPTQOGVKTKATYQSSLLPLQHVKTLLQOOLDALVK 6937

QY 277 ---K-G---N---F--- 280  
Db 6938 IVVARPETHMNEISDQLPISVLSVANSEPOSFYKAGLSLVENHALNNSGGALVFAHD 6997  
QY 281 ---I---LS---AG- 285  
Db 6998 IREGTSMESLTYGBLNTTRANQLANYLISQAKRDELICVCMKESVLSLSILAAVKAGC 7057  
QY 286 ---VD--- 287  
Db 7058 GYLPLVPETPAAIRQILAEADVKFLTDSSMAPVIADVSRCHIMNVOTDTCASQSGCTGP 7117  
QY 288 ---K---T---I--- 291  
Db 7118 QLDFTPTDIAYAVFTSGTTPKPKGVLTQENTLSNLVLSKIYPPVEGSRLLQACNQAFD 7177  
QY 292 ---I-W---D-AH--- 296  
Db 7178 VSVFEIFFTWTYTMCLCSASKDWMFRDPEKAINLEITHLSLTPTVAALTDPAHIPRVKF 7237  
QY 297 -T-GEA---K-Q--- 302  
Db 7238 LVTAGEAVTHVHGAWAGKGLYQGYGPSETTNICTVNSAVESDHDVNNIGPAPENTSAPV 7297  
QY 303 -Q-F-P---F---H---SA---P 310  
Db 7298 LTQGDGFLVPLGGLGELCFGGQVFRGYQNMPELTESKIIHNPVNGRIYRSGDLGRLLP 7357  
QY 311 ---AL---D---V---D---W--- 316  
Db 7358 DGTILIQGRDTPQKIRGORIELGBISGCLLPSPVQONCAIEVIKTADKERLMAFWIPSG 7417  
QY 317 ---Q---S---N--- 319  
Db 7418 YSKDSYSILQDPKNLEEIKSITYAHADNLPAWMPDALVPVSAIPQTSQKIDKRRLAS 7477  
QY 320 ---N---T-FA--- 323  
Db 7478 DGSALTVEDLNAYRGADDDTSELSATEQQLASALADTLQMSQTSIGRSTFFALGLDS 7537  
QY 324 --- 323  
Db 7538 VSAIRLATNLKREYGSIDVSQILKRPTIARLAPLLGGESSKQTNPEPVTADCEAAVGSYL 7597  
QY 324 ---S-C---S---T---DM---C- 330  
Db 7598 HDESUVSOLHEHGQTVSQVLPTQOEAWLSARDTSGSSAYRNKTLFSLHGSVDKUKACW 7657  
QY 331 -I-H---V---C-K---L---G--- 337  
Db 7658 EYMLQRHDLIRTFILSTEDSRFPVQAVLSQWTLPMQECDDIPDQLSTLLDSAKAGDSI 7717  
QY 338 ---Q---DR---PI-K--- 343  
Db 7718 VDHSPPWKIOVYRSESTVYLLDMHLYDANAMSNLLYEVEQLYKQSLASAPVSKFPL 7777  
QY 344 -T---P-Q---G---H--- 348  
Db 7778 NFMISTSVBEADALFRDQLREFVKKPFRKTDVKSFGTITGRNLNYSKPMVETFLSKHSTT 7837  
QY 349 ---T---N---E---V--- 352  
Db 7838 MLSLTQAMWKTLLAASQSYSDVCCGNVSGRSPVVDGIESLVAPCFNTIPVRVDSLSKHSR 7897  
QY 353 N---A---I---K---W--- 357  
Db 7898 NLGLVKALQRVNIDSLPYQLTPLRRIOAQAGTNGKRLFDSLVLLOQDPTDLDLSAIVRLEG 7957  
QY 358 ---D---PT---GNL---L---ASC---SD--- 369  
Db 7958 ESGWMDFPCCIVELAPTNESYTLSLHFNRYLDDDEVVSNLHQACLSAFASCIIRYPSSDVSD 8017

```
QY 370 -----DM---TLK-----I-----WS-MK-Q----- 380
Db 8018 FIDFDADLVAGVLKPDTKHQMPVEAAKTKNRSEKSGSGDESWPLBLQIRAAYSAVSSAP 8077
QY 381 -----D-----N-----C----- 383
Db 8078 EDRIIRDITTYKLGDISISAIQANLRKDKLLVQASDVMSPCSELASAVQSRSQTPV 8137
QY 384 -----D-----LQ-----OH----- 385
Db 8138 LDERGDFEGDKHYRGAALQSHRIATEKVASVRPCTPLQSGMLSEYTHSDGHQYFNHTF 8197
QY 386 -----D-----LQ-----OH----- 390
Db 8198 YAIADIDSKLQSAWSKVLEQHELLRTGFGVDDHEHFPVMLTYTEFDVIDLEIQASSR 8257
QY 391 -----N-----K-----E-----I----- 394
Db 8258 EGSVVEYSQKASESVKNNLHLPFWRWSLGVEGRQCLOFSAHHAIFDAESLRLIMTDLQ 8317
QY 395 -----Y-----TI-----K-----WS-----P----- 401
Db 8318 SALSNGYVTRLTIDGALGHILNSQADVESQRTFWSQKLSGAPVTRFPNMTFVRISDTE 8377
QY 402 -----Y-----L-----LA-S-----A-S----- 403
Db 8378 AANVELVLNKRKLEARQELGVSQMSQGAARLLSAYTGESQVTFGVLSGRTPSA 8437
QY 404 -----P-GTN-----N-----P-NA-N----- 412
Db 8438 TADAAPPCITTLFVSTNTAVDDSQFLKDLMSYNATIQKHQFTPLTIRNYAESTSEALPD 8497
QY 413 -L-----M-----LA-S-----A-S----- 419
Db 8498 SLFVYQRPNDVVDSSWKIIRKASVELAVSVEALSEDGLGLRLTTDPAQVPHQEQK 8557
QY 420 -----F-D-----S-----T-----V-----R----- 425
Db 8558 IMLQMEVMVAGLLKFEADINTSVMSIIPKDPFIATDFKYLHEMTEASVKSYSDDRIAME 8617
QY 426 -----L-----W-----D-----VDRG----- 432
Db 8618 FVDAEDGQISSRHWTYRLQDEANKIAHLIDRGVKGPGDIATSPDKCPEASFAFYGIL 8677
QY 433 -----L-----W-----D-----VDRG----- 433
Db 8678 KAGCAFCADPTAPARKAFILEDNSARVLLTSDTSRSELRBTQCDIIDLINFENKNEL 8737
QY 434 -----C-I-----C-I----- 435
Db 8738 STSSVPVSLGPSVSVLYTSGTTGTPKGCEITHDNVQLVMSFKRLFKGRWTDSERWL 8797
QY 436 -----H-----T-----LTK----- 440
Db 8798 QFASYHFDVSVLEQFTWIVGMELVCAPRDLIEDIAGFLDTMQITHLDTLTPSLGRLLDP 8857
QY 441 -----H-----Q-----E-P----- 444
Db 8858 ALVPSLHKGVFTYTGESLQDQNTWGDVGCPLFNFGPTCTIGVTVPFCVPEKGPSPNI 8917
QY 445 -----V-Y-----S-V-----A-FSP-D-G-R 455
Db 8918 GWQFNVGCVVLAPGQTQPVLRGAIGELCISGLKGVKGYLNRPDLTADCF-PYLDAGFR 8976
QY 456 -Y-----LA-----SGS-F-----D-----K-----C-V-H----- 467
Db 8977 VYRTGDLVRLFDHSGIDFLGRKDNQVLRQORLEIDEBAVIRKQDQIDQTVICVAKHPK 9036
QY 468 -----IW-N-----T-----Q-----VC-----L-----H----- 476
Db 9037 QDKDQLIAFIAGINESKQKPELCPAESTRHLLIQTARAAACEERLPGYVWTFHFLPQIRIP 9096
QY 477 -----Y-L-----N----- 479
```

```
Db 9097 LSVNNKVEKLLRQLYADLPFTTVIQTATQADSOQSLSDGEQKVAQALAEILLKIDNDLT 9156
QY 480 -----G-----QV--LL-N-----L-----GRS----- 489
Db 9157 PSSNIFSLGLSSISAIQFSKSLKASGFTTVQVATVVLKNTPTISRLTKALATSTGRSGGETA 9216
QY 490 -----I-----C-----LY----- 493
Db 9217 DAKQVISACRQRHMGTVTRVLRCKADDIEAIAPCTPLQOGIIISRSLSAESSLYFNSPKFN 9276
QY 494 -----TL----- 495
Db 9277 AQGVLDQLKEGAFNOALERTQILRTFFIETDDGYVQAVRKTGHLPMWLTLEVVDLASVDGV 9336
QY 496 -----P-----HH-----L----- 499
Db 9337 FAKRKQKWRSYNTSHLTVPFEIVIVRSGETFVSDLHHALYDGNSTFDILMNNVSKLYNS 9396
QY 500 -----V-----V-----I-----PL-----V-A-L-I 508
Db 9397 QEADFGKPFVDCLAFGLRNVOGAKQFWDHLDHLPDVKSASMPPLIDNPASHDVLCSTASLDI 9456
QY 509 -----EL-----LV-----L-K 514
Db 9457 LMQADELRRLSLGVTVOALVQATWVATLRK 9485

RESULT 5
I38344
N:Alternate names: connectin
N:Contains: serine/threonine-specific protein kinase (EC 2.7.1.1-)
C:Species: Homo sapiens (man)
C>Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 09-Jul-2004
C:Accession: I38344; I38345; S20898; S20899; S63665; S37393
R:Labelit, S.; Gaute, M.; Laakey, A.; Trinick, J.
S:Science 270, 293-296, 1995
A:Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330; PMID:7569978
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
A:Accession: I38344
A:Molecule type: mRNA
A:Residues: 1-26926 <LAB1>
A:Cross-references: UNIPROT:Q10466; EMBL:X90568; NID:g1017424; PID:g1017425
R:Musco, G.; Tziatzios, C.; Schuck, P.; Pastore, A.
Biochemistry 34, 553-561, 1995
A:Title: Dissecting titin into its structural motifs: identification of an alpha-helix m
A:Reference number: I38345; MUID:95119041; PMID:7819249
A:Accession: I38345
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1977-2014 <MUS>
A:Cross-references: EMBL:X83270; NID:9602579; PIDN:CAAS8243.1; PID:g602580
A:Note: conformation and properties are reported for a synthetic peptide corresponding t
R:Labelit, S.; Gaute, M.; Laakey, A.; Trinick, J.
EMBO J. 11, 1711-1716, 1992
A:Title: Towards a molecular understanding of titin.
A:Reference number: S20897; MUID:92258380; PMID:1582406
A:Accession: S20898
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 13597-14200, '1', 14202-14696 <LAB2>
A:Cross-references: EMBL:X64698; NID:g37192; PIDN:CAA45939.1; PID:g37193
A:Accession: S20897
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 16330-16382, 'S', 16384-16756, 'F', 16758-16860 <LAB3>
A:Cross-references: EMBL:X64699; NID:g37190; PIDN:CAA45940.1; PID:g37191
A:Accession: S20899
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: mRNA
A:Residues: 'P', 22278-22431, 'R', 22433-22448, 'G', 22450-22453, 'Q', 22455-22480, 'TR', 22483-2;
```

A;Cross-references: EMBL:X64697; NID:g37190; PIDN:CAA45938.1; PID:g37195  
R;Kolmerer, B.; Olivieri, N.; Witt, C.C.; Herrmann, B.G.; Labeit, S.  
J. Mol. Biol. 256, 556-563, 1996  
A;Title: Genomic organization of M line titin and its tissue-specific expression in two  
A;Reference number: S63665; MUID:96177761; PMID:8604138  
A;Accession: S63665  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 26729-26825 <KOL>  
A;Cross-references: EMBL:X92412; NID:g1236761  
R;Gautel, M.; Leonard, K.; Labeit, S.  
EMBO J. 12, 3827-3834, 1993  
A;Title: Phosphorylation of KSP motifs in the C-terminal region of titin in differentia  
A;Reference number: S37393; MUID:94008990; PMID:8404852  
A;Accession: S37393  
A;Molecule type: mRNA  
A;Residues: 26831-26926 <GAU>  
R;Improtta, S.; Politou, A.S.; Pastore, A.  
submitted to the Brookhaven Protein Data Bank, February 1996  
A;Reference number: A66736; PDB:1IIT  
A;Contents: annotation; conformation by (1)H-NMR, residues 5253-5341  
R;Pfuhl, M.; Pastore, A.  
submitted to the Brookhaven Protein Data Bank, August 1996  
A;Reference number: A66201; PDB:1NCT  
A;Contents: annotation; conformation by (1)H-NMR, residues 'S', 26059-26155  
C;Genetics:  
A;Gene: GDB:TTN  
A;Cross-references: GDB:127867; OMIM:188840  
A;Map position: 2q31-2q32  
C;Function:  
A;Description: structural protein forming filaments in striated muscle  
C;Superfamily: titin; fibronectin type III repeat homology; immunoglobulin homology; pro  
C;Keywords: alternative splicing; calmodulin binding; cardiac muscle; duplication; glyco  
structural protein  
F;24752-25008/Domain: protein kinase homology <KIN>  
F;84,177,905,2276,2378,2459,2481,2563,2669,2763,2896,3088,3179,3384,3432,3628,3772,4068,  
98,11066,11488,11515,11635,11949,12170,12478,12526,12645,12875,13001,13036,13295,13540,1  
tatus predicted  
F;16780,16976,17579,17602,17667,17681,17845,17899,18121,18188,18209,18336,18670,18680,18  
21900,21935,22295,22495,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248  
F;26171,26178,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 70.6%; Score 2575.8; DB 1; Length 26926;  
Best Local Similarity 8.8%; Pred No. 3.7e-51;  
Matches 441; Conservative 57; Mismatches 11; Indels 4482; Gaps 379;  
QY 1 M-----SISSD-----EV---N-----F-----L----- 11  
DB 20024 MRENTETAEINLTKESVTADAGRYEITAAANSSTGTTKAFINIVLDRPPTGPWIS 20083  
QY 12 -----V-----Y-----R---Y---L-QE-----S-----G----- 20  
DB 20084 DITEESVTLKWPCKYDGGSQVTNYLLKRETSATVWTEVSATVARTMKWKLTTGEY 20143  
QY 21 -----F-----N-----S-----H-----SA 25  
DB 20144 QPRIKAENRGISDHIDSACVTVKLPYTPPGPSTPWTNVTRESITVGWHEPVNNGGSA 20203  
QY 26 -----F-----T-----F-----GI-K-SH----- 33  
DB 20204 VVGYLEMKDRNSILWQKANKLVIRTHFKVTTISAGLIVEFRVYAENAAVGKPSHPSE 20263  
QY 34 ----- 33  
DB 20264 PVLADACEPRNRVITDISKNSVLSWQAPAFDGGSKITGYIVERDLDPGRWTKASFT 20323  
QY 34 -----IS---Q-S-----N---I-N-----G----- 41  
DB 20324 NVTETQFTISGLTQSQYEFVRVARNAGVSNPSEVVGPICTIDSYGGVIDLPLEYTE 20383  
QY 42 -----A-----L---V---P-P-----A-----LI----- 50  
DB 20384 VKYRAGTSVKLRAGISGKPAFTIEWKODKEIQTNALVCVENTTDLASILIKDADRLNS 20443

QY 51 -----S-I-IQ--K-----G-----L-Q----- 58  
DB 20444 GCYELKLRNAMAASATIRVQILDKPFGPGPIEFKVTVAETKITLLWRPPADGGAKITH 20503  
QY 59 Y-VEA-E-----VS-----I-NE-----D----- 69  
DB 20504 YIVEKRETSRVVMSVSEHLESCIITTTKIIKNEIYFRVRAVNYKIGIEPLESDSVAK 20563  
QY 70 -----G-----T-----L-F----- 73  
DB 20564 NAFVTPGPGIEVTKITKNSMTVMSRDIADGSDISGVFLEKRDKSLGWPKVLKETI 20623  
QY 74 -D-----G-----R-----PI-E-S-LSL-----ID-----A----- 86  
DB 20624 RDTROKVTGLTENSQYRVCVNAAGQPFSEPEF-YKAAADPIDPPGPPAKIRIADST 20682  
QY 87 --V-----MP--D-----VW--Q-----T----- 94  
DB 20683 KKSITLGNKSPYVDGSAVTVGVVEIROGEEBEWTVSTKGEVRTTEYVVSNLKPGVNY 20742  
QY 95 -R-----Q-----QA----- 98  
DB 20743 FRVSAVNCAGQGEPIEMNEPVOAKDILEAPEIDLDVALRTSVIAKAGEDVQVLIPFKGRP 20802  
QY 99 -----YR-D-K-L--A-----Q----- 105  
DB 20803 PPTVTRKDEKNLGSARYSIENTOSSLLTIPQVTRNDTGKYLTIENGVGEPKSTVS 20862  
QY 106 -----Q-QA-----AA----- 110  
DB 20863 VKVLDTPAACQKLVQKHSRGITVLLDPPLIDGSGPIINYVIEKRDATKRTWSVVSHKC 20922  
QY 111 -----A-----AA----- 114  
DB 20923 SSTSKLIDLSEKTPFFRVLAEINEIGEGPCETTEPVKAAEVPAPIRDLSMKDKSTKS 20982  
QY 115 -----AA--A-----ASQ--Q----- 121  
DB 20983 ILSWTKPDPDGGSVITEYVVERKGEGTQWSHAGISKTCEIEVSQKQESVLEFRVFAKN 21042  
QY 122 --G-S-----AK-----N----- 126  
DB 21043 EKGLSDPTVITGKILITPEVLDSDIPGAQVTVRIGHNVHLELPYKGPSPISWLK 21102  
QY 127 -G-----EN-T-----A-----N----- 132  
DB 21103 DGLPLKSEBFVRFSTENKITLSIKNAKKEHGGKVTVILDNAVCRIVAPITVITLGPSPK 21162  
QY 133 --G-----E-----E-NG-----AHT-----TA----- 142  
DB 21163 PKGPTRFDEIKADSVILSWDVPEDNGGGEITCYIEKRETSQTNMKNWCVSARTTKVP 21222  
QY 143 N-----N-----H-----T-D-M----- 148  
DB 21223 NLVKDAEYQFVRAENRYGVSQLVSSIIIVAKHQFRIPOPGPKPYVNYVTSQMSLTWDA 21282  
QY 149 -----ME----- 150  
DB 21283 PVYDGGSEVTGPHVEKERNLSILWQKNTSPISGREYRATGLVEGLDYGQFVYAENSAGL 21342  
QY 151 -----VD-G-----DV-E-I-----PB-----NK- 161  
DB 21343 SSPSPDKFTLAVSPVDPGTPDYIDVTRETITLKNPPLRDGGSKIVGYSIEKQGNR 21402  
QY 162 -----AV-----V----- 164  
DB 21403 WTRCNFTDVSECOYTVTGLSPGDRYEFRIIARNAVGTISPPSQSSGIIMTRDENVPPIVE 21462  
QY 165 -----LR-----G----- 167  
DB 21463 FGPEYFDGLIIKSGESLRIKALVQGRPVPRVTFWKDGVIEIKRMNMEITNLVILGSTSLFVR 21522

QY 168 -H-----E-----S-----E-----V-----F-----I----- 174  
Db 21523 DATRDHGVVTVAKNASAKAEIKVKVODTPGKVVGPPIRFTNITGEXMTLWMDAPLND 21582  
QY 175 -C-----AW-----N----- 178  
Db 21583 GCAPITHYIEIKRETSRLAWALIEDCEAQSYTAIKLINGEYQFRVSAVNKFGVGRPLD 21642  
QY 179 --PV-----SD-----L----- 183  
Db 21643 SDPVVAQIYTVDPAGCPEPSNITGNSITLTWARPESDGSIOQYILEREKKSSTRWV 21702  
QY 184 -----L-----ASG-S-----G----- 189  
Db 21703 KVISKRPISRETRFKVGLTEGNEBEYFHVMAENAGVGPASGISRLIKREPVNPPGPTV 21762  
QY 190 -D-S-T-----AR-----I-----W-----N-----LS----- 199  
Db 21763 VKVTDTSKTTVSLWESKVPDVGMEIIGYIEMCKTDLGDWHKVNABACVKTRYTVTDLQ 21822  
QY 200 -E-----N-----S-----T-----Q-----L----- 209  
Db 21823 AGEYKFRVSAINGACKGDSCEVT-GTIKAVDLTAPELDIDANFKQTHVVRAGASIRLP 21881  
QY 210 -----V-----LR-----H-----C----- 214  
Db 21882 IAVQGRPTTAVWSKPDNSLRADIIHTDTSFSTLTVCNNDACKYTLTVENNSGSKS 21941  
QY 215 -----I-----R-----EGG----- 219  
Db 21942 ITFTVKVLDTPGPGPIITFKDVTGRSATLMDAPLLDGGARIHHYVVEKREASRRSQWV 22001  
QY 220 -Q-----D-----VP-----S-----NK-----DV--TS 230  
Db 22002 SEKCTQIIFKVNDLAEGVPYFRVSAVNEYGVGEPEYEMPEPIVATEQAPPRRLDVVDS 22061  
QY 231 -----L-----D-----W----- 233  
Db 22062 KSSAVLAWLKPDDHGSRTGYLLEMRQKSDLWVEAGHTKQLTFTVERLVEKTEYEFV 22121  
QY 234 -N-----SE-----G-T--LL--A-----T-----G----- 243  
Db 22122 KAKNDAGSEPREAFSIIKEPQIEPTADLTGINTQLITCKAGSPFTIDVPISGRPAK 22181  
QY 244 -SY-----D-----G-----F-----A----- 249  
Db 22182 VTWKEEMRLKETDRVSIITTKDRTTLTVKDSMRGDSGRYFLTENTAGTKTFSVTVVVI 22241  
QY 250 -R-----I-----W--TKDG-----NL-----ASTLG----- 263  
Db 22242 GRPGVTGPIEVSSVSAESCVLSWGEPKDGCGTEITNYIVEKRES--GTAMQLVNSSVK 22299  
QY 264 -Q-----H-----K-----G-----PI-----F-----A 271  
Db 22300 RTQIKVTLTKYMEYSFRVSSNRFVSKPLSAPIIAEBHPFVPPSAPTRPEVYHSANA 22359  
QY 272 -----L-----K-----W-----NK----- 276  
Db 22360 MSIRWEEPYHDGSKIIGYWEKKEKERTILWKENKVPCLCNYKVTGLVGLGYQRTY 22419  
QY 277 ----- 276  
Db 22420 ALNAGVSKASEASRPIMANQPNVDAPRPEVTDVTRSTVSLIWSAPAYDGGKVKVYIIE 22479  
QY 277 -K-----G-----NF--I--LS-----AGV----- 286  
Db 22480 RKPVSVEGDRWLKCNVTIVSDNFTVTALSEGDTVEFRVLAKNAAGVLSKGSSTGPT 22539  
QY 287 -----DK----- 288  
Db 22540 CRDEYAPPAELDARLHGLVTRAGSDLVDAAVGGKPEPKIITWTKGDKELDLCEKVL 22599  
QY 289 --T--T--I----- 292

Db 22600 QYTKRATAVIKFCDRSDSGKYTLTVKNASGTKAVSMVKVLDSPGCGKLTVSRVTQEK 22659  
QY 293 -W-----D--A--H-----T-----GE----- 299  
Db 22660 CTLAWSLPDGDGAETHYIVERRETSLRNWVIVEGECPTLSVYVTRLIKKNYIFRVRA 22719  
QY 300 -----A-----K-----Q----- 302  
Db 22720 VNKGPGVPESEPIVARNSTFTIPSPGPIEEVGTGCKEHIITQWTKPESDGGNEISNLYV 22779  
QY 303 -----QF-----P-----F----- 306  
Db 22780 DKREKESLWTRVNKDYVYVYDTRLKVTSLMEGCDYQFRVTAVNAAGNSPERSNFISCR 22839  
QY 307 -----HS--A--PALD-----V-----D-W-Q--SN-- 319  
Db 22840 BPSYTPGSPSAPRVVDTTKHSISLAWTKPMYDGGTDIVGVLEMQSKOTDQWYRVHTNAT 22899  
QY 320 --NT-----F----- 322  
Db 22900 IRNTEFTVPDLKMGQKYSFRVAANVKNMSEYSESIAETIEPVERIEIPDLELADDLKTV 22959  
QY 323 -----AS----- 324  
Db 22960 TIRAGASLRMLVSVGRPPPVITWSKQIDLASRAIIDTTESYSLLIVDKWRYDAGKYT 23019  
QY 325 -----C-----S-----T-D----- 328  
Db 23020 IEAENQSGKKSATVLVKVYDTPGCPSPVKVKEVSRDSVTITWEIPTDGAPINNYIVEK 23079  
QY 329 -M-----C-----I-----H----- 332  
Db 23080 REAMRAFKYTVTKGSKTLIRISGLVEGTMHFRVLPENIYGIGEPCESTDVAVLSEVPL 23139  
QY 333 -----V-----CK----- 335  
Db 23140 VPAKLEVVDTKSTVTLAWEKPLYDGGSLRTGVLEACKAGTERMKNVTKPTVLEHTV 23199  
QY 336 -L-----G-----QD-R--P-I-----KT----- 344  
Db 23200 TSLNEGEQYLFRIRAQNEKGVSEPRETVTAVTVDLRLVPTIDLSTMPQKTIHPVAGRPV 23259  
QY 345 -----FOG-----HT-----N- 350  
Db 23260 ELVPIAGPPPAASWFFAGSKLRSESVTVETHTKVAKLTIRETTIRDTGTGEYTLKKNV 23319  
QY 351 -----EVNA-----IKW-----D----- 358  
Db 23320 TGTSETIKVIILDKGPGPTGPIKIDEIDATISITISWEPPELDGGAPLSGYVVEQDAHR 23379  
QY 359 -P-----T-GN-----LLAS-----C--S----- 368  
Db 23380 PGWLPVSESVTRSTFKFTRLTEGNEVFRVAATNRFGIGSYLQSEVIECKRSSIRIPGPE 23439  
QY 369 -D--D-MTL-----K-----I-----W----- 376  
Db 23440 TLQIFDVSRRDGMTLTWYPEDDGGSQVTGYIVERKEVRADRWVRVNVKVPVTWTRYESTGL 23499  
QY 377 -----S-----M-----K-Q--Q--DN----- 382  
Db 23500 TEGLEYEHRVTAINARGSKPSRPSKPIVAMDPIAPPGKQPNPRVTDTRTSVSLAWSVP 23559  
QY 383 -----C-----V-----H----- 385  
Db 23560 EDEGSKVTGYLIEQKVDQHEWTKCNTTPTKIREVTLTHLPGQABYRFRVLACNAGGPG 23619  
QY 386 -----D--L-----Q-----Q- 389  
Db 23620 EPAEVPGTVKVTEMLEYDPYELDERYQEGIFVRQGGVIRLTIPIKKGPPPICKWTKEGOD 23679  
QY 390 -----H-----N--KE-IYTIK----- 398

```
Db 23680 ISKRAMIATSETHTELVIKEADRGDSGTVDLVLENKCGKAVY-IKRVIGSSENSPEGPL 23738
QY 399 -----WSP-----TG-----P-----CT-----N 407
Db 23739 EYDDIQRSVRVSRPADDGGADILGILYLERREVPAKAWYTIDSRVGTSLVVKGLKEN 23798
QY 408 -----N-----P-N-----AN-----L----- 413
Db 23799 VEYHFRVSAENQGISKPLKSEBPVTKPLNPPPPSPNPPEVLDVTKSSVLSWSRPRKD 23858
QY 414 -----M-----L-----A----- 416
Db 23859 DGGSRVTGYIERKETSTDKVVRHNKTQITMTYVTGLVPAEYQFRIIAQNDVGLSET 23918
QY 417 S-AS-----F-----DS-T-----V-R----- 425
Db 23919 SPASEPVVCKDPKPSQPGELEILSISKDSVTLQWEKPCDGGKEILGYWVEYROSQDS 23978
QY 426 LW----- 427
Db 23979 ANKKSNERIKDKQFTIGGLEATEYEFVPAENETGLSRPRTAMSIKTKLTSGEAPGI 24038
QY 428 -----DV-----D-R-G-I----- 433
Db 24039 RKEMKDVTTKLGAAQLSCQIVGRPLPDIKWYRFGKELIQSRKYKMSDGRTHLTVMTE 24098
QY 434 -----CI-----H-----TL----- 438
Db 24099 EQDEGVYTCIATNEVEVETSKLLQATPQHPGYLKEKYKYGAVGSTLRHLHWYIGR 24158
QY 439 -----SV-----A-P-----SP----- 442
Db 24159 PVPAMTFHCKQLQNSENITIENTHYTHVMKNVQRKTHAGKYKVLQSNVPGTVDAIL 24218
QY 443 -----E-----P-----V-----Y----- 446
Db 24219 DVEIQDKPKPTGPVIEALLKNSAVISWKPADDGGSWITNVVVEKCEKAEGAEMQLVS 24278
QY 447 -----SV-----A-P-----SP----- 452
Db 24279 SAISVTTCTRVNLTENAGYFVRSANTFGISDPLEVSSVVIKSPFEPGAPGKTITA 24338
QY 453 -----DG-----R-YL-----A-S-----G----- 460
Db 24339 VTKDCSVWAKPPASDGGAKIRNYLLEKREKKONKWSITTEIRETVFSVKNLISGLEV 24398
QY 461 SF-----D-----K-----C----- 465
Db 24399 EPRVKCNLGESESEISEPITPKSDVPIQAPHPKEELRLNVRYSNATLVCKVTGHP 24458
QY 466 ----- 465
Db 24459 KPIVKWYRQKEIIADGLKYRIQEPFGYGHQLIIASVTDDATVYQVRATNQGGSVSGTA 24518
QY 466 -----VH-----N----- 470
Db 24519 SLEVEVPAKTHLPKTLLEGMAVHALREVVS IKIPSGKPDPIVTQKQDLDLNNGHQY 24578
QY 471 -----T-----Q----- 472
Db 24579 VIVTRSFSLVFPNGVERKDAGFYVCAKNRFGIDQKTVELDVADVPDPGRGVKVSASGR 24638
QY 473 -----V-C----- 474
Db 24639 DSVNLWTPEPASDGSKITNYIVEKATTAERWLVRVQARETRYVINLFGKTSYQFVII 24698
QY 475 -----L-----H-----Y----- 477
Db 24699 AENKFGLSKPESEPIITKEDKTRAMNYDEVDRETSMTKASHSTIKELYKMYIAE 24758
QY 478 -----L-----NG-OVL-----LN----- 485
Db 24759 DLGRGEFGIVHRCVETSCKKTYMAKFKVKGTDQVLVKKEISILNIAHRNHLHLSPE 24818
```

```
QY 486 -----L-----G-----R-S-----I-----C-----L----- 492
Db 24819 SMEELWMIFEFISGLDIFERINTSAPFELNEREIVSVHVQVCEALQFLHSHNIGHDPDIRPE 24878
QY 493 -----Y-----T-----L-----P-H-H----- 498
Db 24879 NIYYTRRSSTIKITEFGQARQLKPGDNFRLLFTAPEYVAPEVHQHDVVVSTATDMWSLGT 24938
QY 499 LV-V-----I-P-L-----V-----AL-----IE-----LLV----- 512
Db 24939 LVYVLLSGINPFLAETNOQIIENIMNAEYTFDEEAFKEISIEAMDFVDRLLVKKRKRMT 24998
QY 513 -----LK 514
Db 24999 ASEALQHPWLK 25009

RESULT 6
T42737
gp330 protein precursor - rat
N:Alternate names: megalin
C:Species: Rattus norvegicus (Norway rat)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 09-Jul-2004
C:Accession: T42737
R:Saito, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994
A:Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of the
A:Reference number: A58173; MUID:95024033; PMID:7937880
A:Accession: T42737
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-4660 <S>
A:Cross-references: UNIPROT:P98158; EMBL:L34049; NID:G561852; PID:G561853; PIDN:AAA51369.
A:Experimental source: strain Sprague-Dawley; kidney
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
F:1-25/Domain: signal sequence #status predicted <SIG>
F:26-4660/Product: gp330 protein #status predicted <MAT>

Query Match 70.6%; Score 2575.1; DB 2; Length 4660;
Best Local Similarity 10.1%; Pred. No. 9.6e-53;
Matches 422; Conservative 62; Mismatches 27; Indels 3679; Gaps 367;

QY 1 M-----S-----I-S-----SDEVN----- 9
Db 11 MLLLAIAACLEPVSQEGSGNFRCDNGYCI PASWRCGTRDCLDDTDEIGCPPRSCSG 70
QY 10 -FL-----V-----YR----- 14
Db 71 LFLCPAEGTCIPSPSWVCDRDKDCSDGADPQNCAGTTCSAQQMTCNSGQCIPSEYRCDHV 130
QY 15 -----Y-----L-Q-----E- 18
Db 131 SDCPGSDBERNCHYPTCDQLTCAAGACYNYSORCDQKVDRCRSSDEANCTTLCSSQKEFEC 190
QY 19 -SG-----F-SH-----S-AF-T-----F-G-I----- 30
Db 191 GSGECILRAYCHDNDCEDNEDNERNYDTCGGHQTCSNGQCINQNVWVCDGDDDCQDS 250
QY 31 -----K-SH-----IS-----QSN----- 39
Db 251 GDEGCGESQSHRVCYPREWACPGSGRCISDKVCDGVDPCEPGEEDENNVTSGRTCMGV 310
QY 40 -----N-----GALV-PPAALI-----S-----I-I-OK----- 55
Db 311 CSVLNCEYCHOTPPGEGCFPPGHIINSNDSRTCIDFDQCIWIGICDQKCNQRQHQC 370
QY 56 -----G-L-Q----- 58
Db 371 LCEGYILRGQHKSSDSFSAASVFSNGRDLVLGDLHGRNFRILAESKNRGMVMGVDF 430
QY 59 -Y-----V-----E-----AE-VS-----IN-----E----- 68
```

Db 431 HYQKRVFWTDMQEKVFSTDIINGLNTQBIILNSVDTPENLAVDMINNKLYLVETKVRNI 490  
QY 69 D---G---TL---F-DG-R--- 76  
Db 491 DVVNLEGNQRVTLITENLGHPRGIALDPTVGYLFFSDWGLSQCPKVERAFMDGSRKDL 550  
QY 77 ---P---IE---SL--- 83  
Db 551 VTTKVGWPAGITLDLVSKRVYVDSRYDIETVTDYGIQKTVARGGSLVPHFGISLFE 610  
QY 84 --I--D--AVM---PDVV-Q---T---RQ--- 96  
Db 611 EHVFTDWTMAVMKASKFTETNPQVYHOSLSRPHGVTVYHALRQPNATNPGCSNNGCA 670  
QY 97 QA---YR---D---K---L---A---Q--- 105  
Db 671 QVCVLSHRDNGGLGRCKCEFGFELDDDEHRCVAVKNFLLFSSKTAVRGIPFTLSTQED 730  
QY 106 ---Q---Q---Q---A--- 108  
Db 731 VMVPVTGSPSFFVGIDFDAQHSTVFYDLSKDIYKQKIDGTGKEVITANRLESVECLTF 790  
QY 109 ---A---A---AA---A--- 112  
Db 791 DMISRLYWTGGLKSVTVLRLADKSRQIISLNNPRSIIVHPTAGYMFSLSDWFRPAKI 850  
QY 113 --A--A---A---AAA---A---S---Q--- 120  
Db 851 MRAWGSHLMPVINTSLGWPNGLADWSASRLYWDAPFDKIEHSTLDGLRKLGHVD 910  
QY 121 Q---G-SA-K-N---GE 128  
Db 911 QMTHPFGLTVFXDNVFTDWRLGAIIRVRKSGDGMTVIRRGISSVMHVKAVDADIQTGS 970  
QY 129 N---T-ANG---E---E--- 135  
Db 971 NYCSTTHANGDGHCFPPVPNQRVCGPYGMKLQDQMTCEGDPAREPPTQQCGSLSF 1030  
QY 136 ---NG---AH---T---I---AN 143  
Db 1031 PCNNGKCVSPFRCDGVDDCHDNSDEHQCVFNNTCSPAFACVRGQCIPGWHCDRON 1090  
QY 144 ---N---H-T---D-M---ME-V---DG-DVE--- 156  
Db 1091 DCLDGSEQNCPTHATSTCPSFTCDNHVCIPKDWCDTNDSCDSD-EKNCOASGT 1149  
QY 157 ---I--P---P---N---K-A---V--- 163  
Db 1150 COPTQPRCPDHRICISPLYVCDGDKDCADGSDGAGCVLNTCTSAQFKCADGSSCINSRYCD 1209  
QY 164 -VL---R-G-H---E---S-E---V- 172  
Db 1210 GVTDCRDNDGACPRPFGMCHPDBFCQGDGTCTPNTWECGHPDCIHGSDHTGCVP 1269  
QY 173 ---FIC---AM---N---P--- 179  
Db 1270 KTCSPHFLCDNGNCYKAWICDGDNDCRDMSDEKDCPTOPHCPSTQWQCPCYSTCINL 1329  
QY 180 ---V---S---D---L---LA--- 185  
Db 1330 SALCDGVFCPNGTDESPLCNQDSCSHFGGCTHQCWQGPFGATCLCPLGYQLANDTKC 1389  
QY 186 ---S---GS---G--- 189  
Db 1390 EDINECDIFGCSQHCVNMRGSPRCADPEYLTLESGRCTKVTSSENPLLWASRDKIIV 1449  
QY 190 ---DS-TARI---W---N---L 198  
Db 1450 DNTATHNLYSLVQDVSFVALDFUSVTCRVPWSDLLQKTSVSPQNGTDKRVVHDSGL 1509  
QY 199 S-E---N---S---T---S---GS---T---Q-LVL-R--- 212  
Db 1510 SVTEMIADWIGRNLWYVALETIEBVSXKIDGSHRTVLISKVTKPRGLADPRMGDNVM 1569

QY 213 ---H---C--- 215  
Db 1570 FMSDWGHHPRIERASMDGTMTRTVIQEKIYWPCLSIDYPNRLIYFMDAYLDYIEFCDYD 1629  
QY 216 ---R---E---GG--- 219  
Db 1630 GHNRRQVIASDLVLRHHPHALTLFEDFVYWTDRGTROVMQANKWHGQNSVVMYSHVQPLG 1689  
QY 220 ---Q---D---V---P--- 223  
Db 1690 ITAIHPSRPPSRNPNPCASCSHLLCSAQAPRHYSACPSGNLSDSDSVNCRGQDPFL 1749  
QY 224 -S--N---K-D-V-TS---LD---W-N-SE---G 237  
Db 1750 MSVRDNIIFGISLDPEVKSNDAMVPISGIGHGVDFEFDDEQFIYVVENPGEIHRVKTDG 1809  
QY 238 ---T---L-LA---T---G-S-Y---D--- 246  
Db 1810 SNRTVPAPLSLGLSSGLALDWMVSRNIYVTPASRSIEVLTILKGDTRYGKTLIANDGTPL 1869  
QY 247 --GF---AR-I-W---T---K---D---GNI--- 258  
Db 1870 GVGFPVGIADVPAKGLYMSDHTGDSGPAPKIASANMDGTSKILFTGNLQHLLEVTLDI 1929  
QY 259 ---A-S---T---LG---Q--- 264  
Db 1930 QEQKLYWATSRGVIERGNVDTGTERMILVHHLAHPGLVYGSFLYYSDEQYIEVERDK 1989  
QY 265 ---HK---G---P---I--- 269  
Db 1990 SSGNKNVLRDNPVYLRGLRVYHRRNAADSSNGCSNPNACQICLPVPGMFSCACASG 2049  
QY 270 ---P---A---LK--- 273  
Db 2050 FKLSPDGRSCSPYNSFMVVMPLPAVRGFSLELSHSEAMVPVAGQGRNVLHADVDVANGF 2109  
QY 274 --W---N---KK-G-NF---I---L-SAG--- 285  
Db 2110 IYWCDFSSSRSSNGIRRIKPDGNSFTNVVYIGANGIRGVALDWAAGNLYFTNAPVE 2169  
QY 286 ---VD---K---T---T--- 290  
Db 2170 TLIEVLINTYVRRVLLKVSVDMPRHIIVDPKHYLFWADYGQKPKIERSFLDCTNRTVL 2229  
QY 291 ---I---I-W---D-A-H---T---GE 299  
Db 2230 VSEGIPTPRGLAMHDHTGYIYVWDSLDLIARIHLDDGESQVVRYSRYPTPYGITVFE 2289  
QY 300 ---A-KQ---Q--- 303  
Db 2290 SIIWDRNLKKVQASKQPGNTDPPVIRDKINLLRDVTIFDEHAQPLSPAELNNPCLQ 2349  
QY 304 ---F---P---F---H 307  
Db 2350 SNGGCSHFCEALPELTPRCGCAFGTLGNDGKSCATSBQEDFLIYSLNLSRLHFDPRDH 2409  
QY 308 SAP---ALD---V- 314  
Db 2410 SLFPQVISVAGTAIALDYDRNRNRIFFTOKLSLRQISVSYLSGSSPTVLLSNIGVT 2469  
QY 315 ---DW---Q---SN--- 319  
Db 2470 DGIADFWRNRIYSDFSNQITNSMAEDGSRNAVIRVSKPRAIPLDPCRGYMYTDMGT 2529  
QY 320 ---NT--- 321  
Db 2530 NAKIERATLGGNFRPVIIVNTSLVWPNGLALDLETLLYWADASLQIERSTLTGTNRVV 2589  
QY 322 ---P---A---S--- 324  
Db 2590 VSTAFHSFGLTVYGOYIYWTDLTRYKRIYRANKYDGSDLVAMTTRLTQPQSGISTVVKTQR 2649

QY 325 ---CST-D-M-C-I---H---V---CK---L 336  
Db 2650 QCCSNPCDFNGGSHICAPGNAECQCPEGWYLANDNKYCVDTGTRCQLQFTCL 2709  
QY 337 -G---QD---R-P---I---K---T---F--- 339  
Db 2710 NGHCINQDKNDNDGDSDELPTVCAFHTRCSTAFTCGNGRCVPHYRYCDYNDGCD 2769  
QY 340 ---R---P---I---K---T---F--- 345  
Db 2770 NSDEAGCLFRNCNSTTFTCSNGRCIPLSVVCNGINCHNDTSDERKCPPTCPDFTK 2829  
QY 346 -Q---G---HT---NEVNA---IK-W--- 357  
Db 2830 CQTNICVPRAFLCDGNDGSDENPIYCASHTCRSNEF-QCLSPQRCIPSWFCDGE 2888  
QY 358 ---D-P-T-G---NLL-AS---C-S---D---DM--- 371  
Db 2889 ADCADGSDPTCGHVSNTCRASQFQCDNGRCISGNWVCDGNDGDSDEDRHHCCLQ 2948  
QY 372 ---T-L---K-I-W---S---M--- 378  
Db 2949 NCSSTQTCVNSRPNRRRCIPQYVWVCDGADCDALDELQNMRTCSAGEFSCANGRCV 3008  
QY 379 KQ---D-N---C-V-H---DLO--- 388  
Db 3009 RQSPRCDRNDGDSYDERGCSYPPCHANQFTCNGRCIPRFVFCDEDNDCGDSDEQEH 3068  
QY 389 ---Q---HN--- 391  
Db 3069 LCHTPEPTCLHQPCDNGHCIEGRVNCNVHVDGSDNSDEKGCINECLDSISRCDHNC 3128  
QY 392 ---KE--- 393  
Db 3129 TDTITSFYCCLPGYKLMDSKRSVDIDECKESPOLCSQKCNVNVGSYICKAPGYREP 3188  
QY 394 ---I---Y--- 395  
Db 3189 DGKSCRQNSIEPYLIFSRYIRNLTTDSSYSLLIQLGNGVVALDFDREKELWIDA 3248  
QY 396 ---TI---K-W---S--- 400  
Db 3249 EKQIIRMFNLKNTRETIINHLRRAESLAVDWVSRKLYLWLDAILDCLFVSDLEGRHKM 3308  
QY 401 ---PTG---P---GTN---N--- 408  
Db 3309 IAQHCVDANNFCFEHPRGIVLHPQRGVYVWADGWVHAYIGRIGMDGTNKSVIISTKIEW 3368  
QY 409 PNA---N-LML-ASA---SF---D-S---TVRLW--- 427  
Db 3369 PNAITIDYTNLLYWADAHLYIEFSLEGGHRRHTVVDGSLPHFPALTIFEDTV-FWTDW 3427  
QY 428 ---D---V---D---R---G---I--- 433  
Db 3428 NTRTEKNGKYDGSRVLVNTTKPFDIHVYHPYQPIMSNPGTNGGCSHLCLIKAG 3487  
QY 434 ---CI--- 435  
Db 3488 GRGFTACPDPTQVQLRDLTCLWPMCSSTQFLCGNNEKCIPIWKCQKQCSGSDGDEP 3547  
QY 436 ---H---T---L---T---KH--- 441  
Db 3548 DLCPHPRCLGQFCRDGNCSTSPQALCNARQDCADGSDDEDRVLCHEHRCSNEWQCANR 3607  
QY 442 ---Q---E---P---V---YS--- 447  
Db 3608 CIPQSWQCSVNDCLONSDEDTSHCASRTCPGQFKCNGRCIPQSWKCDVNDGCDYSD 3667  
QY 448 ---V---A---FS---P---D---G--- 454  
Db 3668 EPIDECTAAYCNDNHTFESCKTNYRCIPQWAVCNGFDCCDNDSDGQCSVPCPSGDF 3727  
QY 455 ---RY---L-A---S---G 460

Db 3728 RCANHHCIPLRWKCDGTDGCDNSDENVCVPRCSESEFRCADQOQIPSRWVCDQNDG 3787  
QY 461 ---S---F---DK---C--- 465  
Db 3788 DNSDERDCMKTCHEPHFOCTSGHCVKALACDGRADCLDASDESACPTFRFPNGTYCPAA 3847  
QY 466 -V---H-I-W---NT---Q--- 472  
Db 3848 MFECKNHVICQSWICGENDCVDSDEBIHLFCFNIPCESPQRCFRCDNSRCVYGHQLCNG 3907  
QY 473 V---C-LH---Y-L-NG---Q---V---L---LNLG 487  
Db 3908 VDCGDSDEKEEHCRCRPHKPCCTDEYKCSNGNCISQHYVCDVNDVDCGLSDSETCNLG 3967  
QY 488 ---RS---IC---L--- 492  
Db 3968 DNRCTAENICEQNCCTQLSSGGFICSCRPKGPSTSKNSQDINECEEFGICPQSCRNSK 4027  
QY 493 ---Y---T---LP---H--- 497  
Db 4028 GSYECFCVDFGFKSMSTHYGERCAADGSPPLLLPENVRIRKYNTSEKSEYLEEHEHIQ 4087  
QY 498 ---H-L-VV---I---LV---L---K 514  
Db 4088 TIDYDWDPEHIGLSVYVTVLAQGSQFGAICRAYIPNFESSGNNPIREVDLGLKYLQMPD 4147  
QY 506 ---A---L---IE---L--- 514  
Db 4148 GLAVDWGRHYWSDAKSQRIEATLDGRYKWLITQLDQPAIAVNEPK 4197

RESULT 7  
T15789  
hypothetical protein C41A3.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T15789  
R;Bentley, D.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of C. elegans cosmid C41A3.  
A;Reference number: Z18404  
A;Accession: T15789  
A;Status: preliminary; translated from GB/EMBL/DBBJ  
A;Molecule type: DNA  
A;Residues: 1-7829 <BEN>  
A;Cross-references: UNIPROT:Q18559; EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA8311  
C;Genetics:  
A;Gene: CRSP:C41A3.1  
A;Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/2;  
/2; 7514/3; 7596/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3  
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein  
F;6547-6616/Domain: acyl carrier protein homology <ACP>  
F;2832,5271,6580/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 70.5%; Score 2574.2; DB 2; Length 7829;  
Best Local Similarity 7.9%; Pred.No. 3e-52; Indels 5238; Gaps 381;  
Matches 457; Conservative 46; Mismatches 10; Indels 5238; Gaps 381;

QY 1 MS---ISS---D---E---V-N---F---LV--- 12  
Db 1 MSACLVGWSWAKTPNGEDGHEMAENIFLNRNNTALLPVEKNYLSDFREKHPEKVALVDGI 60  
QY 13 ---Y---R---Y- 15  
Db 61 EYFDDQYFGTGESEAIACMDPQOQMLMQGVIKGLGNAGITLEMASEARVAVYTAAMCYDK 120  
QY 16 ---L---Q---E---SG--- 20  
Db 121 DLLPDDQWATGNSASVMCGRITYFLNSRGAAGVETACSSSLVAFHARQAIQSGETKL 180  
QY 21 ---F---SH---SA-F---TF--- 28



Db 181 ALVCGANHVGRSFBHSLYNSHMVSPNGRLAAFDRSANGFVRAESFAVAVLCSKQFAEENN 240  
QY 29 -----G-----I-K-S-----H----- 33  
Db 241 LLHCECVSAFNSDGTPLSTAPNPISQYEVQLEALKNIDKDSVQLVTCHGTGTKLGDQ 300  
QY 34 -----I-----S-----OS-----N----- 38  
Db 301 VELTAINRSPKSDIRVMSPKSMGSHGGAAGLIGVLQSLYSMHQGIIPNQLHLELPSED 360  
QY 39 -----IN-----G-----ALV--P--P-----A-----A-- 48  
Db 361 GEDKSMGFVNEEMELNRVAISSYFGGTNACATIEKPEKPSLVQKESYAESNVFLSAKS 420  
QY 49 -----L-I-----S-----I-----I-----I-----QK----- 56  
Db 421 HESLKQIEEYTFMAQDSAMEDILYTVNERKTYDFRAAVFGKDNEIARKLQGDYS 480  
QY 57 -----LQ-----Y-----V-----EA-----E----- 63  
Db 481 LTNLQBSTFEVEFEGNEKMLLRLMYEKNETFHSVDKYCKLAETCGPPEATALFFPF 540  
QY 64 -----VS-I-----N----- 67  
Db 541 KLTLELTYNVSRLISSMATFELLVQYNTLPNKLKGLQIFCLAVAKVITFESAVOLI 600  
QY 68 -----E-----D-----G-----T----- 71  
Db 601 KGVVAEANLTDLGDIELKSSKIPIBIQHILKSTKKILPIHISGELKETAKPNLMTPIVN 660  
QY 72 -----L-----FD-----GR-----P-----I 78  
Db 661 GBEILELDPVRKVOKLICQLFACGFPDPAVKFRGRIVKTPTYSFLKQFPWQVQTAMTNI 720  
QY 79 -----E-----SLS-----L-----I-----D-----AV----- 87  
Db 721 VDEQTNSSLSDAEISTVRTIVKQFLDIBEDDINLLETGAVDLSLTIEMVEAFGTAVNQT 780  
QY 88 MP-----T-----R-----Q-----D--VV-----Q----- 93  
Db 781 MPFDLLEAYPTLINIVDFLKLTVTPVTKATTSIHKKTSELSTSDINVIACDYQPAGVE 840  
QY 94 -----T-----R-----Q----- 96  
Db 841 GEKELWTLTSLRTLTKGIDIRKKQCEGDAGLEVGLLKQDISMPDNSSFFATAKDEAFL 900  
QY 97 -----Q-----AY----- 99  
Db 901 DPQHRLLNAAYNALKEKSLTIPDADFLAISAHSEYRALAEKHINELDERLMMGTVHS 960  
QY 100 -----R--D-----KLA-----Q--Q----- 106  
Db 961 MVAGRLAVLMGIRGRAMIVDTTCSSVATALEMAVKSIREGRPAIVATSQLIOSKWLVS 1020  
QY 107 -----Q-----A-----A-----A-----A----- 111  
Db 1021 LKTLDDHSTNSFSDGSGFRSDGVVILKTAEGDSAVIKISAKSHHCCAVMTVPV 1080  
QY 112 -----A-----A-----A-----A-----A-----AS----- 119  
Db 1081 SSISQLLEBAGFSYVEGHGTATSDAGDSAESMAYQKLGSELIMSSVKAQFGHCEVASGLI 1140  
QY 120 Q-----Q----- 122  
Db 1141 QLMKVSSIGKHGIIPISVHNILPSEHNRNENIRLPFVAEEKQIDRSATVSFGITGKTV 1200  
QY 123 -----SAK----- 125  
Db 1201 VITERVSQNLVDNIQNCYLLPVSATKTDGKAKCLSLIEMIDNSCESLYDISTTLOKQKT 1260  
QY 126 N-----G-----E-NT--AN-----G-----E----- 134  
Db 1261 NPKWRTAVVSGSHADVVLKQFLTSEHNTSLTNWHISTSHHSIGCSTFFHNIPEFDHY 1320

QY 135 -----E-----NG-----A-----H-T-----IA-N----- 143  
Db 1321 SMFCHRLRPEPHSNTNSIYHLLAVVYALIRVILKHKLKLTNSFAVGGFNSLIVLAADAAPS 1380  
QY 144 N--HT--D--MM-----E-----V 151  
Db 1381 HYLNDLLHAPANDVRRMMKRIARDVTISTENVKLLNLNGEPIITTARQAVEATIDQKVXNV 1440  
QY 152 -----D-----G--D-----V--E-- 156  
Db 1441 RLPETTLILSPRAVEFASQLETIQDYKYLIGEKTQGGQVDPAGIFGTPIKULIDPEYP 1500  
QY 157 -----IP--P--N-----KA-----VVL-----R- 166  
Db 1501 FNRKSFWLPIPDVSPSNEKEKPLIPKSYEFLLKSQKQWQHVHVDSKIVLPGATSIRL 1560  
QY 167 -----G----- 167  
Db 1561 VHQLANGKPTVELSNIDFLNKITPSEAPSVVKIEEQDLEKLVFGETDAISFKLTQLQNFN 1620  
QY 168 -----H-----E-----S-----EV-F----- 173  
Db 1621 PIPNERLNAEVHHTNIYERFANSHLTYRNEFQMDVSLKYTMKGGEVRFVSMKDLID 1680  
QY 174 -----I--C----- 175  
Db 1681 GTLQALVGCYFFENTNDNSPFVFTIDQLSILNGDISQQLHAVLYKVDSSGNPINGDATV 1740  
QY 176 -----A----- 176  
Db 1741 YDALGNIILHISNVTFKRLNGQASPLSTKTVDSKSTKTKVENEQDKRASKMHLWVFE 1800  
QY 177 ---W-----N-P-----V-----S-DILASG----- 187  
Db 1801 NFGWTDIDNTGFFDLGLTSIOAVKLRNAKSNYPNASSTCVFDPYPSIDLL-SGYLSTLN 1859  
QY 188 -----S-G-----D-S-----T-----A-R-----I-----W----- 196  
Db 1860 DPQVTSIGEDDIOKDLTEDHKPFLAENPIGVMAACRLPGGVSSPSSELWELLKIGN 1919  
QY 197 -----N-L-----S----- 199  
Db 1920 ASSRIPATRPVTRNTLISGSKYGNPVEGGNFITQDVTQDPDFSKSEALIDPQQL 1979  
QY 200 -----ENS--TS-----GS----- 206  
Db 1980 LLECVOECLSENSGVITETSNVGVFGLMEKBYQDMMESSILAMLGSMMAAIVAGRVNYIFG 2039  
QY 207 -----T----- 208  
Db 2040 CYGPSVTIDTACSSSLVALEMAINALLDNRCVKIVAGVNLILNEKGQGLRTNGKLSQH 2099  
QY 209 ---L-----VL-----R----- 212  
Db 2100 GMSLSPDSRASGYGRSDGCVLMLELAKPNFHMSTIQSVNVNHHGGRSVSLTAPNVAHK 2159  
QY 213 -----H----- 213  
Db 2160 MLTTSVINQSPSLAIDYWEAHGTGTPGLDPIEBFTLSSILQNIIGSVKASLGHGEASAG 2219  
QY 214 -C-----IR-----E-----G--G-- 219  
Db 2220 TCGLLKLFLMLTYQVPTLIHFHVLNKOINAGSIRLPIIGEDSELVSAGISFGVSGTNA 2279  
QY 220 -----Q-----DVP----- 223  
Db 2280 AAIAPNDNNKLPYPIHKYYILPISAKNQISLDNLEKQILSVIPLTDVPICNIASALAN 2339  
QY 224 -----SN-----KD----- 227  
Db 2340 NRSHTFIRNALIVNSGIVNSKMEGPHRAKDRYHVKLCDLSDLLQYDVINETYT 2399

QY 228 V-----TSI-----D-----W--N-S-E-- 236  
Db 2400 VASLKNQSFAMKFAIKFSLSEVIEIVASDGBELLAVLANGSLKWFNFKTWIELP 2459  
QY 237 -GTLL--A-----TG-----S 244  
Db 2460 IGSLLTEFADHLNLTSSIKSYQTPSSHNLDSPELMKLMKIYITGYDWDATVYS 2519  
QY 245 -----Y-----D--G-----F-- 248  
Db 2520 PVEQFALPNYQFNKQTLWFERLEIVDHYLIGTIDESEDTLILKNQISELHPQFPKG 2579  
QY 249 -----A--R-----I--W--T----- 253  
Db 2580 KPLDVCTMSEIAIEALKIRNEIPFSIONLKTETILTTPAWLETNVTNRNEDDEGFNVSA 2639  
QY 254 -----K----- 254  
Db 2640 IDQRLFSLNASSVEIQIEVPAVEQIPDKVVYLKECPNAVIRRRNMVYVDSRAEQSP 2699  
QY 255 -----DG-----N-- 257  
Db 2700 FRTANIVLNEIIGFAPTPSDFEILGVLPVSRHYMVQVDDGALWQPMISQDKRVLNIIY 2759  
QY 258 -----LA----- 259  
Db 2760 VLKDAKGLPIPTIRMHKSTLLSSQBSASIVAATLQMAVRHKVCLAVGDVIESGLDIDES 2819  
QY 260 --ST--LG----- 263  
Db 2820 QLSGTSELGIDSLATVDLLNRLNKQYFPEIELTTSDFNPSIIDLSIMIBQLLNEKGI 2879  
QY 264 -----O-----H-- 265  
Db 2880 TEPSEPNTPKTSLRGRKLSIPAVRAQVLAQIEFVENYNSKQEVQAEAPSSSECSNHL 2939  
QY 266 -----KG-----P 268  
Db 2940 ESDATVDRTEIRKVS LAVFDLATETLSAEDLSQKGFTELGMDSLSIVDFVRLNDKYFP 2999  
QY 269 -----IF-----A----- 271  
Db 3000 DDEITASDFDYPTVDELSDHIVRKSSSVPPAASEIMKETMNGISTSVDAEHTKLENLS 3059  
QY 272 -----LK--W--N--K--K--CNFI--LSA-----G----- 285  
Db 3060 QSPMLLENQNSINPTLKMINSQTIKLVKPSDGNFLFELNANGQGEKEIQKHFTGPNII 3119  
QY 286 VD-K-----T--T-----I----- 291  
Db 3120 IDLKGFEHGSTETLYMSLLNLVKSISKLEIQCFGVSGFGLGNSISRAFMKTVAEKNP 3179  
QY 292 -I--W-----D----- 294  
Db 3180 LISFAWQNVQVSVFVDSPIGTGNLITGGLSGIGLEIGKFIANGNAENVILISRRQPT 3239  
QY 295 --A-----HTG----- 298  
Db 3240 AKALRDILSTELTHIGLARTIVILKIKNISAKLIVQSFKLSFSFKYISDLFSKKVTF 3299  
QY 299 -----E----- 299  
Db 3300 FYNLQSCFCKHFWLFEWLEVYLKQWTVHTIAADINDKEKLIBELTKLVNGITGIH 3359  
QY 300 -A--K--Q-----Q--P--P--FH----- 307  
Db 3360 SAGVLDKSKIERQKESFNQVTPKANGFHVLEIEKHFNYKIENFIMMSFTAACNEG 3419  
QY 308 -----S----- 308  
Db 3420 QLNYSVNAYLEYOVORRRRQSGCAIQGNWIDTGMATDENVRKFLANLGLFOHNDK 3479  
QY 309 -----A-----PAL-----DVDW-----Q-----S 318

Db 3480 ALKYLRACILTKPELIMIVANIDMNVILKNRDLPKDLINTGILPFDFTGKNESEFPDLS 3539  
QY 319 N----- 319  
Db 3540 NGDFEKVSMNFSVEDEEVELELIKVKSSILCMSPTKLKNKNKIMDMGLDSKLIIVEFLNF 3599  
QY 320 -N--TP-----A-----S--C----- 325  
Db 3600 INSTFKISVNLSDAYNHPTLEKLAHIFEQMTIVDHPVNSVKSEIFKSTDCPIFGINI 3659  
QY 326 -----ST-----D----- 328  
Db 3660 PFDKNDFDKAKSTAVKLENGEQLPTAGKYAVSVVGKSIIRDVWSKIEAAPQIKLQCE 3719  
QY 329 -----M-----C-----I----- 331  
Db 3720 SSKCVMLTGOGSQYPMGRQLVENEYIFRTTLOSLCKKDEYLOQDVSLEWILFNTHY 3779  
QY 332 -----H--V--C----- 334  
Db 3780 KLLQTKMQPIMFCGYATAQLMISLGIVPDYILGSHVSELVAGVLAGIMSIEDGLRLI 3839  
QY 335 ---K-----LG---Q---D---R---P--I---KT-----F-- 345  
Db 3840 VERGKAMENIAGLGALLAVQREIADVLKRPKVSATINSPKQVVPAGTGSVLDAAAEV 3899  
QY 346 --QG-----HTN--E-----VN---A 354  
Db 3900 KQGGKQATVYVNOQYPPHNSLIQETHLVSLRQCLADIKFSAGRTPLVSNVTGOIINTFSEA 3959  
QY 355 -I-----K--W--D-----PT----- 360  
Db 3960 YIVKHTVSAKVEQVDCVETLOAKGVTVWIDAGSAVLATFVKRIIQPTLSKRIIVQTCKE 4019  
QY 361 -----GNLL-A-----S----- 368  
Db 4020 KESDNDVLVQACLELEQSGLPISWTLYGCGRNADERLVEFPMTHTNDIINKDEFELLEGH 4079  
QY 369 -----D-----DMT-- 372  
Db 4080 QLNGKIVVAGAYQLFKIDQLVKLKAAGMELMLKNVKFLKPWYIEDNREYQIOWNSDMTIE 4139  
QY 373 -----LK----- 374  
Db 4140 LIVNSVIVCSLEVEPQNSVLKLETISENEKEPVEHDFYETLFRNGILOYDSGFRRIESARR 4199  
QY 375 -----I-----W-----S--MK----- 379  
Db 4200 SDRKCFSQIKSSPFAWPLIDSAMHSITASVVRPRDCYFLPVAMGSVMTKDTNFTLPNL 4259  
QY 380 --Q-----D--N-----C-----VH----- 385  
Db 4260 HAQTVITSETDKFIQNVNALLAGDTPICEVRNMTIVLTKTEPVTIRIPNSIETVETPPKS 4319  
QY 386 -----D--L--O-----OH-----N--KE--I----- 394  
Db 4320 EIEIVGFDISLPYQI SENSENWOHLKNTYVKQKLHNSLKQDHARVALLDSADRYWDEP 4379  
QY 395 Y-----T-----I-----K 398  
Db 4380 YFGIRPSEAKFTDPOORLLCSVAKLLDSLLITSNTSGVFIGCSANBFHIVVAYGK 4439  
QY 399 -----WS-----P-----TG----- 403  
Db 4440 DPAEWSGTSNSALAGRIAHWLKLGVPVVTLDTACSSSFYALSAACDALRTGQCEYAI 4499  
QY 404 -----P----- 404  
Db 4500 GTVNLVMEHTTVDLQNAKQVTDDFCFADFVDANGYKREAVCSMLLTKSPNIDSVATIT 4559  
QY 405 -----GT-----N-----NP-----NA----- 411

Db 4560 NYATGHNGTSSLTFTNGLSQLEVMQORATNPLEKILEIOTHCTGTCLGDPFIEINAIKSLV 4619  
QY 412 -----N-----LM-----K 414  
Db 4620 SSACKIGSVKSNIGHTEGSGLVSLCSSLMSFRSKYRVAQLHLKCPNTSNKTKWMCIRFI 4679  
QY 415 -----L-----A-----SA-S----- 419  
Db 4680 GEDADENNSILINNFPGTSGNCSWLKPKNAISEHFVSSEVFPYILLSSHSAKSLQKYVQ 4739  
QY 420 -----D-----S-----T-----VR----- 425  
Db 4740 VLCEFISNSAKSLHIMMSLFQKKIHVHRQFIIPNFKRIAVTSLDGFVEVRDERLEKX 4799  
QY 426 -----L-W-----D-VDR----- 431  
Db 4800 HPCSFLKEGVVHFDKDKSFQVDLPISIVNNTLHWALDSYRDEIDRHESQMSFKNIFY 4859  
QY 432 -----GI----- 433  
Db 4860 EKVLETPMPQONPIKVVVICGRDLIPKIEIDVSSFANGIIVPHPISNSIFEYLKLA 4919  
QY 434 -----CI-----HT-----L-----T-----K-----H----- 441  
Db 4920 VMSLISRNQNVFIICCFENGTSHTMTGTLRSLASEKMIPIYKFVSDIKVDALKLJEFNHE 4979  
QY 442 -----Q----- 445  
Db 4980 YMFELFYKRGYVERLKVNFTEKPAQYKELISGGTGGIGSAIINELKPKSSVIIT 5039  
QY 446 -----Y----- 446  
Db 5040 RKNIASDGTPLSSDITRLDISHKENVYFHLAGIVNNSLHENVKRDSDLDEWVSIKLOGA 5099  
QY 447 -----S-V-----AFS-----P-D----- 453  
Db 5100 KNLKCCDETHFVFSSIANVLGYSQSNYAFSGLVTSFLETSTKSTIHWGPWKDV 5159  
QY 454 G-----R-----Y-L-----A-S-----G----- 460  
Db 5160 GMLAQPERRIVKQIESNGKWLPNQDALSIVFTQFMETHQIIVFDGDFDTIVARQPHL 5219  
QY 461 -----S-----F-----D-----K-----C----- 465  
Db 5220 OKLSEVVEKTKYKBEIKKSLNFEEIPEIVGIDTSSKLNIPFMDLGLDSLCMENRY 5279  
QY 466 -----V-----H----- 467  
Db 5280 SLNKNFDLELTVSEMFENATYQKQTYVETLRAKHSLSLSDRVSSQVSNKEDDTRVAV 5339  
QY 468 I-W-----NT-----Q-----VC-----L----- 475  
Db 5340 IGWSAEFGSSNIHEYENLMDGICSTGNKYLKKNPFGDNKFFNLTDARVLPQVR 5399  
QY 476 -----H-YL-----N-G-----Q----- 481  
Db 5400 KFIQHAYLALENSYYKQHELRCGVFAGAPESDYGRADDDHDAEKLFWMNNSYLASY 5459  
QY 482 -----VL-----L-----N-----L----- 486  
Db 5460 ASYCLDLKGEAVSVYSGACSTALVAVANAVKSIQSGSMYALVGAASIAEVSGALSGFDDQ 5519  
QY 487 -----G-R-----S-I-----C-L-----Y-T-----L----- 495  
Db 5520 KKTMFSGVCRPFDKDSGIVRGSGVGVFLKRYSQLLDNDNVHFVKDFAINNDGHS 5579  
QY 496 -----P-----H----- 497  
Db 5580 RASFMAPAGQLKCMTDVLARFTNKEKERSIFVECHATGTTLGTIEMNSLRTAYSFKN 5639  
QY 498 -----H-----LV-----VIP----- 503  
Db 5640 KLAIGSKANIGHAYAAAGLAALVKCAKMLQGTIIPQVNFSEFRDGMGQFFTVNGKST 5699

QY 504 -----LVAL-----IEL-----LVL-----K 514  
Db 5700 ISQNSLISIDSPGIGGTNNVHMVIEFPARQSEVVVKISSENILLYDMIPISAK 5750

## RESULT 8

Tl3931

projectin - fruit fly (Drosophila melanogaster) (fragment)

C:Species: Drosophila melanogaster

C&gt;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C:Accession: Tl3931

R:Daley, J.; Southgate, R.; Ayme-Southgate, A.

J. Mol. Biol. 279, 201-210, 1998

A:Title: Structure of the Drosophila projectin protein: isoforms and implication for pro-

A:Reference number: Z17815; MUID:98300339; PMID:9636710

A:Accession: Tl3931

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-6658 &lt;DAL&gt;

A:Cross-references: UNIPROT:O76281; EMBL:AF047475; NID:g33337430; PID:g33337431; PIDN:AAC3-

A:Gene: projectin

A:Cross-references: FlyBase:FBgn0005666

A:Map position: 4

A&gt;Note: intron positions not resolved (incomplete sequence)

C:Keywords: muscle

Query Match 70.5%; Score 2572.1; DB 2; Length 6658;

Best Local Similarity 9.1%; Pred. No. 2.4e-52;

Matches 438; Conservative 66; Mismatches 5; Indels 4309; Gaps 375;

QY 1 MSI-----S-----S-----D-----E-----V-----N-----F----- 10

Db 20 KSLIAKRFVKPFVGEGLKNTVVKGTIRFDIKYDGEPEPAATVWKTDLNLFQNR 79

QY 11 -----L-----V-----Y-----L-----Q-----ESG-F----- 21

Db 80 CLDQLERNSSITIKSVKDKTGKYLVLNSSGTIESEAQVVVLDRLPLPGGPFEEIR 139

QY 22 -SH-----S-----A-----F-----T-----F 28

Db 140 ASHIKMKWKRPDDGGCEISGVALERMDEETGRWIPAGEVGNPNETSFDFKGLTPNKKYK 199

QY 29 -----Q-----I-----KS----- 32

Db 200 RVKAINKEGESPLETFDASCNPPYDPPSPQPVIDDYDNKSVLLKWKRPSPDCGRPT 259

QY 33 H-----I-----SQ----- 36

Db 260 HYIVEIKDKFAPSWSEVAKTDDPNPECNVEGLKEMVYQFRVAVNKNAGSPSPQTDNH 319

QY 37 -----S----- 37

Db 320 LCKHKNLKPDIDRSTFKRVTIKSGRTHKWSVDLGEPIELHWSWRDDIPLTNGDRIKIE 379

QY 38 N-----I-----NG-----A-----A----- 42

Db 380 NVDYHTDFSITNVLKRDGSGFYTLKAENRNGIDRETVELVLGKPSPKGLAVSDVTASG 439

QY 43 --L-----V-----PP----- 46

Db 440 SKLQWKKPDEVGVPIKEVVEKMDTATCKVVRVGRSPGKEKPPFDVTGLSLGSEYMPR 499

QY 47 --A-----LI-----SI-----IQ 54

Db 500 VSAVNEGDAEPLTTLVGVAKDPFDPNKPCTPEVTDYDNQISLSLKAAPNNDGCAPIQ 559

QY 55 K-----GLQ----- 58

Db 560 KYIIEKKNKNTWEKALEIPGDQLEATVAGLQYGEYQFRVIANKAGLSPSPDASVPQ 619

QY 59 -----Y-----V-----E-A-----E 63



QY 297 -----T-GE-----AK-Q-----Q----- 303  
Db 2838 VGKLDETTGRWMTAGETDGVNTALKVGGTLPGHKYKFRVRAKNRQGTSEBLTTAQAIK 2897  
QY 304 -----F-----P-----F----- 306  
Db 2898 NPPDVPTKPTPTIKDFKDFVLEWTRPEADGGSPITGVVVEKDKFSPDMEKCAEISD 2957  
QY 307 -----H-----S-----A-----P----- 310  
Db 2958 DITIAHVPLIEGLKYEFVRVAVNAKAGPSPSDATETHVAPKNTPPKIDRNPMSEDIK 3017  
QY 311 A-----LDV-----DW-QSN-NT-----F----- 322  
Db 3018 AGNVFEDVPVTGEPLPSKDWTHEGNMIINTDRVKISNFDRTKIRILSATSOTGYTILT 3077  
QY 323 -----A-S-C-----S-T----- 327  
Db 3078 ARNINGTDRHNKVTILDAPSVEPALRNGDVSKNSIVLRWRPPKDDGGSEITHYVVEKM 3137  
QY 328 D-M-----C-----I-H-----VCK-----L----- 336  
Db 3138 DNEAMRWVPGDCTDEIRADNLNENHDYSFRVAVNAKQSQPLTTSQPIAKDPYSHP 3197  
QY 337 -----GO-----D-----R-----PI-----KTP-Q-----G-----H 348  
Db 3198 DKPGQOATDWGKHFDLEWSTPKRDGGAPISYIIEKRPK-FGQWERAADVLDGNCKAH 3256  
QY 349 -----TV-----E-----VN-----AI-K----- 356  
Db 3257 VPELTNGGEYEFVRVAVNAKQSQPLTTSQPIAKDPYSHP 3316  
QY 357 W-----D-----P-T-----G-N-----L-L----- 364  
Db 3317 WTLPIASPRPLTILWYNGKEIGSNRSGSLFQNELTTEIVSSLSRSGRYTLILKNEH 3376  
QY 365 -----AS-----S-----C-----S-----D----- 369  
Db 3377 GSFDAHAHATVDRPSPKGLDITKTRDGGCHLTWNPVDDGGSPILHYIEKMDLSRS 3436  
QY 370 -----D-M-----T-----L-----L-----K 374  
Db 3437 TWSADAGMTHIVHDVTRLVHRKEYLFRVAVNAIGESDPLEAVNTIIAKNEFEDPDAPGK 3496  
QY 375 -----I-W-----S-----MK----- 379  
Db 3497 LIITDWRDHIQLQWAVPKSDGAPISEYIIQKKEGSPYTNVVRHVPKNTTIPILT 3556  
QY 380 -----Q-----D-----N-----C-V-H-D----- 386  
Db 3557 EGQVEYFRVAVNAKQSQSPSPDMIMKPRYLPPKIITPLNEVRKICGLIFHTDIHI 3616  
QY 387 -----L-----L-----QO-----H----- 390  
Db 3617 GEPAPATWTLNSPLNSDRSTITSIGHSVVHTVNCQSDSGIVHLLRNSSGIDEGS 3676  
QY 391 -----N-----N-----K-----EI-YT-K-----W----- 399  
Db 3677 FELVLDKPPGEGMEYEIEITANSVTISWKKPPKONGSEISSYVIEKRDLTHTGGGWPA 3736  
QY 400 -----S-----P-----S-P-----T-G-PG----- 405  
Db 3737 VNVSAKNHVAVPRLLGTYELRVMAENLQGRSDPLTSDQPVVAKSQYTVPGAGKPE 3796  
QY 406 -----T-----N-----N-----P----- 409  
Db 3797 LTDSDQKHITIKWQPIISNGGSPGIIIGYDIERRDVTNRKINGQVPVTAEQDDRVTSN 3856  
QY 410 -----NA-----N-----L-M-----L-----A----- 416  
Db 3857 HOYQVIRISAVNAAGNKTSEPSAIFNARPLREKPRFYFDGLIGKRIKVRAGEPVNLNIP 3916

QY 417 S-A-----S-----F-----DS----- 422  
Db 3917 SGAPTPIEWKRGDLKLEBKGRISYBETNSERTLFRIDDSNRDRSGKYTVTAANFCKOTA 3976  
QY 423 -----TV-----R----- 422  
Db 3977 DIEVIVDKPSPPEGLSYTETAPDHISLHWYSPKDDGGSDITGYIIEFTFGVDDMKPV 4036  
QY 423 -----TV-----R----- 425  
Db 4037 PGTCPTNTFTVKNLVEGKKYVFRIRAEINYGASEALEGKPVLVKSPDPGAPSQPTISA 4096  
QY 426 -----L-W-----DVD-----RG-----I-C----- 434  
Db 4097 YTPNSANLEWHPPD-DCGKPKITGYIVERRERGGEWKNNYPTPNTSVTVSNLRDARY 4155  
QY 435 -----I-----H-----T-LT-KHO-----EP-----VYS-----VAFS-----P-DG-454  
Db 4156 EPRVLAVNEAGPHSPKSPDPMTABHORYRPPPEPKPDRI-TRNGVTLSMRPPTDGGK 4214  
QY 455 -R-----YLA-----S-G-----SF-----D----- 463  
Db 4215 SRIKGYVEMRPNGKDKWTVNDIPINVTYVTPSLKEGEEYSFRVVAENEVGRSDPSKP 4274  
QY 464 -----K-----C----- 466  
Db 4275 SQPITIEQPNKPCMELGKVRDIVCRAGDDFSIHVPYLAFFPKPNAPWYNDNMDNNRV 4334  
QY 467 H-----I----- 468  
Db 4335 HKHLTDAASVVVKNKSRDSQYRLQFKDTSQFDTATINVRVLRPSPTLRADVFSG 4394  
QY 469 -----WN-----T-Q-----V-C----- 474  
Db 4395 DSLTYWNPDDGSAQNYIIEKKEARSSTWKSUSSECTVLFVRIRNLVNLKEYDFRV 4454  
QY 475 -----L-----H----- 476  
Db 4455 IAENKYGQSDPANTSEPIILARHPFDIPNTPGPHGIDSTEDSTIIATWKPKHGGSPITG 4514  
QY 477 Y-----L-----N-GO----- 481  
Db 4515 YIIEKRLSDDKWTKAVHALCPDLCKIPNLNIAEYEFRAVNAAGOSAYSGSSDLIF 4574  
QY 482 -----V-L-----L-L-----G----- 487  
Db 4575 CRPPHAPKITSDDLIPDMTVIAGDEFRITVTPYHASPRTASWSLNGLEVIPEERIKFDS 4634  
QY 488 -----R-----S-I-----C-----L-Y----- 493  
Db 4635 NDYASMYNKSARKDETGSYITLTNNKGSDDTASCHVTVDRLPPQGPLNAYDITPDTG 4694  
QY 494 TL-----P-----H-----HL-----V----- 500  
Db 4695 TLAWKTPLDGGSPITNYVVEKLDNSGWSWKISSFVRNTHYDVMGLEPHYKYNFRVAEN 4754  
QY 501 -----VI-PLVA-----L----- 509  
Db 4755 QYGLSDPLDIEFPMVAKHQFTVPDEPGQPKVIDWDSGNVTLIWTRPLSGGSRIOGYQIE 4814  
QY 510 -----L-----LVLK 514  
Db 4815 YRDILNDSWNAYDIK 4832

## RESULT 9

A45086

HC-toxin synthetase - fungus (Cochliobolus carbonum)

C:Species: Cochliobolus carbonum

C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 03-Nov-2000

C:Accession: A45086

R:Scott-Craig, J.S.; Panaccione, D.G.; Pocard, J.A.; Walton, J.D.

J. Biol. Chem. 267, 26044-26049, 1992

A;Title: The cyclic peptide synthetase catalyzing HC-toxin production in the filamentous  
 A;Reference number: A45086; MUID:93100328; PMID:1281482  
 A;Accession: A45086  
 A;Status: preliminary; not compared with conceptual translation  
 A;Molecule type: nucleic acid  
 A;Residues: 1-5232 <SCOs>  
 A;Experimental source: SB111  
 A;Note: sequence extracted from NCBI backbone (NCBIP:120884)  
 C;Superfamily: Cochliobolus carbonum HC-toxin synthetase; acetate-CoA ligase homology; a  
 C;Keywords: carrier protein; phosphopantetheine; phosphoprotein  
 F;277-743/Domain: acetate-CoA ligase homology <AC11>  
 F;784-854/Domain: acyl carrier protein homology <ACP1>  
 F;1882-2366/Domain: acetate-CoA ligase homology <ACL2>  
 F;2397-2464/Domain: acyl carrier protein homology <ACP2>  
 F;3036-3526/Domain: acetate-CoA ligase homology <ACL3>  
 F;3549-3618/Domain: acyl carrier protein homology <ACP3>  
 F;4186-4655/Domain: acetate-CoA ligase homology <ACL4>  
 F;4680-4750/Domain: acyl carrier protein homology <ACP4>  
 F;818,3583,4715/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 70.48; Score 2568.8; DB 2; Length 5232;  
 Best Local Similarity 8.9%; Pred. No. 1.8e-52;  
 Matches 447; Conservative 54; Mismatches 6; Indels 4542; Gaps 379;  
 1 MS-----I-SS--DE--V-NF-----LVY----- 13  
 179 MSAMEKVNVSSTLYIMSSDPDKTAINRLSIPRDLAQIMRWNRDLKKSERTNLVYDLFS 238  
 14 R-Y-----R-Y-----LQ-----E-----S-19  
 239 ARAHEQDANMAIDAWDGRMSYTELERVSTSWARQLOKQIGISQGSWVLFCEKSRLLAVSM 298  
 20 -----G-----F-S-----22  
 299 IAILKAGVCVPIIDPYPVERIRDIIRTNATIALVAGKTAALFKSADTAVOTIDTKD 358  
 23 --H-----S-----AF-TF-----GI-K--SH-----33  
 359 IPHGLSTVVQSNKTDPAFGLFTSGTGVKCIUVVTHSQICTAVQAYKDRFGVTSETR 418  
 34 --IS-----Q-----36  
 419 VLOFSSYTFDISADTFALFYGGTLCIPSEEDRMNLQDMVSVRPNWAVLPTVSRFL 478  
 37 -----SNI 39  
 479 DPGVVDFTLFTGASREADTVPIEAGVNLVNYGPAENTLITATIRKKGKSSNI 538  
 40 --N-----GA-LVP-----P-----AA-----L-I-----50  
 539 GYGNTRTWTVDYSGACLVPGSIGELLIESGHLADKYLNRDTEAFLSDLPWIPNYE 598  
 51 --SI-----I-----Q-K--GLQ---Y--VEA-----E 63  
 599 GDSVRRGRFRRTAGDLVRYCDDGLICVGRSDTQIKLAG-QRVELGDVEAHLQSDPTTSQ 657  
 64 --V---S-----I-----NEDGT-----L--F-----73  
 658 AAVFPRSGPLEARLIALLVGNKDGTPHNQSLPKPAFAQCPPDLVKYATSLQORLPS 717  
 74 -----D-----G--R-----P-----I-----E-----79  
 718 YMVPSVWLIGDPLPMSVSGKLDRAVLQDQLESLSPSDYAEILGTTGLEVDPGCAASSVAS 777  
 80 -----SL-----SLI-----D-----A--V---M-----88  
 778 DSDLRMDNDDSLLTACSRVNLNPAKISYSQSFIHAGGDSITAMQVSSNMKRFCTGKRIG 837  
 89 -----P-----D-----V-V---QTR-----Q-----A-----Y-99  
 838 VKDLLVSPSISTAACTKSAQDSRNFVAPQO-RIPVSPQKLFQTAESKSNWYH 896  
 100 -----R-D--K-----L-AQ-----Q-----AAA-----110

Db 897 QSFLRIDQPIKQPTTIEDAISLMQRHPLMQARPERTBEGDWYQYIPIDVEREASVEVIG 956  
 QY 111 -----AA---A-----A-----A-----115  
 Db 957 SLSTDDREAAMLRARQSIDLTEGPLIRCOLFNNNVDEASRLFPFWIHHAVIDLVSWRIIM 1016  
 QY 116 ---A---A---S-----Q-CGSA-KN-----G 127  
 Db 1017 EELEAHLATDSTPDGRGEAYOESVFLAWCQOABAVKIPVDRTVPLIPKIPTADFGYWG 1076  
 QY 128 ---ENT-AN-----G---E-----E-----135  
 Db 1077 LKHDENVYNTVERKIPGLGHSITEDLLYKCHDSLHTKTDIVLLAAVLVSFRSFLDRPVP 1136  
 QY 136 ---N-GAH-----TI-----A--NN 144  
 Db 1137 AVFNEG-HGREPGGEDAVDLSRTVGWFTTISPVYVEVSPGDILDVVRVKDYRWATPNN 1195  
 QY 145 -----H-----T-----D-147  
 Db 1196 GPDYFTKYLTSQGIKLFEDHLPAEILFNYEGRYQAMESEQTVLKPEWHAGEASKDDDP 1255  
 QY 148 -M---M-E---V--DG-----D-----V-EI-----P 158  
 Db 1256 GLRRFCLFEISTAVILPDGQLHLTCMNKMRHQGRIRLMDTLTPAAIGEVSSLASASP 1315  
 QY 159 -----PNKAV-----VLRG-H-----ES-170  
 Db 1316 QUTLSDVELLRLYDSSLDILKKSILSIP--AVOTLDDLEGVYVPGSPMODALFLOSQSK 1373  
 QY 171 ---EV-F-----I-C---AM-----N-----P-----179  
 Db 1374 DGAYEVDFTWRVATSLQNSQPAVDIGCLVEAKWDTVALHAALRTVILESSLPATGILHQV 1433  
 QY 180 V--S--D--LL-----A-----S-----GS-----188  
 Db 1434 VLRSHPDVIDLDRVDTAAITILDSYPPPTBEGIALIKRPPHRLICTTIEGSLVKFQ 1493  
 QY 189 -----G--D-----S-----T-----A-----193  
 Db 1494 VNHLPDGMSTDKIIQDLSKAYTCRHSNKLDPHSESKLHDGTGYNRPTRKPLAEFIRYR 1553  
 QY 194 --R--I--W-N-----L-----SE-----200  
 Db 1554 DPQRKQDSINYKWNALRGATTCSPPLFDQITSEKAMPQSWASVPIPLCVDSKELSKTL 1613  
 QY 201 -N-----ST-----S--G--S-----T--Q-----LV-----210  
 Db 1614 ANLGITMSTMFQTVMAIVLRIYSONGSQSVFGLTSGRDAPVDGIDSAVGNFIAMLVCFED 1673  
 QY 211 L-----R-----H--C-----IRE--G-----218  
 Db 1674 FDDGVHTVADMARKIHNASANSISHQACSLAEIQDALGLSTSTPLFNFTATYLPKRPTN 1733  
 QY 219 --GO-----D--V--P--SNK-----D-----V--TS-----230  
 Db 1734 VKAGEPEHLCELELSMSDPTFEDLTFLVEPTQESNEVSAHLDFKLSYISQAVATSIAT 1793  
 QY 231 -----L-----D-----WN-----234  
 Db 1794 VAHILSELVHDVPRALNTLPVSEHDTAIRSWNDHLPPPTECHETFSRKVHEHPQRE 1853  
 QY 235 -----S---E-----GT-----238  
 Db 1854 AICSDGSLTYAELSQRLSILHLSVSLGKIVGTIKIPICEKSMWTVITLAVVQAGGVF 1913  
 QY 239 -----LL-----240  
 Db 1914 VLLEPCHPSRLSGIIKQVQAEALLCSPATSRMGALQNLSTQMGTEFKIVELEPEFIRSL 1973  
 QY 241 -----ATG-----243

Db 1974 PLPPKPNHOMVGNDDLYVFTSGTVKPGAVATHQAYATGIYEHAVACGMTSLGAPP 2033  
QY 244 ---SY-D-G-F-A---R---I---W---T--- 253  
Db 2034 RSLQFASYSFASIGDIFTLAVGGCLCPREDRNPAGITTFINRYGVTWAGITPSLAL 2093  
QY 254 ---KD---G---NL---A---ST--- 261  
Db 2094 HLPDPAVPTLKALCVAGEPLMSVWTVWSKRLNLINMYGTEATVACIANQVTCITTTVS 2153  
QY 262 ---LG---Q---H---KG---P---IPALK---W--- 274  
Db 2154 DIGRGYRATVWVQPDHNSLVPFAGVAGELIIEGILCRGLYNDPRTAEVF-IRSPSML 2212  
QY 275 ---N---K--- 276  
Db 2213 HDLRPNSTLYKTGDLVRYSDGKLIIPGRKDTQVKNQGFELGEVEHALQLQDPSDGP 2272  
QY 277 ---K---GN---FI---LS--- 283  
Db 2273 IIVDLLKRTQSGRPDLIIAFLFVGRANTGTGNSDEIFATSTSSLSEFSTVIKKLQDAQR 2332  
QY 284 ---AG-VD--- 288  
Db 2333 AMEVLELFMVPOAYIPIEGGIPLTAAGKIDRRMLKLCBPFNENDLISFTSKALSTSVKD 2392  
QY 289 ---T-TI---IW--- 293  
Db 2393 AETTDVERLARIWEKVLGVKGVBRESDFSSGGNSMAIALRAEQRSGFTLFVADIF 2452  
QY 294 ---D-A---H--- 296  
Db 2453 TNPRLADMAKLFSGHOSVSPSSSTLRKVPISLQKRSSGLOTAAPVNSGSPVRRCKEN 2512  
QY 297 ---T---G---EA--- 300  
Db 2513 IIDCPVAFYEBGSPDTOLKEASRICGISRSIEDVFPCTPQOEALVALSLIPGAQASYA 2572  
QY 301 ---K--- 301  
Db 2573 LHAAPFLRGLDRNRRFSAWESTVKAQPIILRSRIISGNSGVVVTSATDPSIPOLDVSGL 2632  
QY 302 ---QO---F---P---F---H---S---A--- 309  
Db 2633 DTLEQOLQVGFAPGAPFLPRLAFVYSKADDCDFVISAHHAIYDGSNLNLSQVLATY 2692  
QY 310 ---P---A---LD---V-D--- 315  
Db 2693 NGELPPPGPSFKHFAARNLNLVQSKLDSDFWRKLLVKPDQESFRFPDVPVGHKPATRCTT 2752  
QY 316 ---W---Q-SN---N---T---F--- 322  
Db 2753 NFHPFPMOSKIGTTANTCINAAMAITLAQYSSNKTVNFVTLWGRDFPMIDIEHMTGPT 2812  
QY 323 ---A---S---C-ST--- 327  
Db 2813 IIVTPQVNVIPSSVAEFLODLQKSLAVVLPQHLGLHRIQALGPQACDFSLVV 2872  
QY 328 ---D---M---CI--- 331  
Db 2873 NHGSSISWSELEAADIVPPLRSSDLIAYPMVVEVENASSDTLDIRVHSDPDCEVQLLE 2932  
QY 332 ---H---VC--- 335  
Db 2933 RLMEQFNLQTLCLRAASFDPGKRIAEMLDDTATHTLTFWSNSRVKSDPVAIAVHK 2992  
QY 336 ---QO---DR---P-I--- 342  
Db 2993 LLEETAQSPAESAIVAHQDQLSYMQMDRCADVLARQIRKTNMISAQSPFVCIHLRSAT 3052  
QY 343 ---K---TF--- 345  
Db 3053 AVVSMVLAVLKAGGAFMPVDISQPRSLQNLIIEBSGAKLVLTLPESANALATLSGLTKVIP 3112

QY 346 ---Q---G---H---T--- 349  
Db 3113 VLSSELVQOITDNTTKKDEYCKSGDTPSSPAYLLYTSGTSGKPKGVVMEHRAWLSLPTC 3172  
QY 350 ---N---E-VN--- 353  
Db 3173 HAEYMGFNSCTRIFLOPSSLMFOLSIIEIWAIVLYAGGCLFIPSKDKERVNNLQDTRINDIN 3232  
QY 354 ---A-I---K---W--- 357  
Db 3233 TVFLTPSIKLNPKDLPNISFAGFGEPMTRSLIDAWTLPGRRLVNSYGPTEACVLVTA 3292  
QY 358 ---D-PT---G-N---L---A-S---C--- 367  
Db 3293 REISPTAPHDKSSNIGHALGANIWWVEPQRTALVPFAGVAGELCIEAPSLARCYLANPER 3352  
QY 368 --- 367  
Db 3353 TEYSFPSTVLNDNWTKKGTRVYRTGDLVRYASDGLDFLGRKDGOIKLGRQIELGEIEH 3412  
QY 368 ---SDD--- 371  
Db 3413 HIRRLMSDDPRPHEASVQLYNPATDPRDATVDVQMRPYLAGLLVLDVLSMRSDSMH 3472  
QY 372 ---TL--- 373  
Db 3473 VLNIANTSNIQTLVTELKKSRLGVLPHYMVPLHFVAVSRLPTGSSGKLDHAFVACLRE 3532  
QY 374 ---KI---W--- 376  
Db 3533 LTAPLDGNPFKEQVLTNTNESVLRQWGTVLAMDPHSIQGRDDFFSLGSSISAMRLVGL 3592  
QY 377 ---S---M---K--- 379  
Db 3593 ARSSGHKLQHEDIFMCPRLADMAGQISFVQEASVSPTTPTKFDLLDCEVDEVIDHIL 3652  
QY 380 ---Q-D-N---C--- 383  
Db 3653 PQLDNNKELLIEDVYPTCPLQESLMAATARHGEAYTMIQSIIVLAAQLAKKAMDVVFRD 3712  
QY 384 ---V-H---D---L---Q--- 388  
Db 3713 FEVLRTRIALGPSQALQVWVKHELSWESFPISQSKDHFYRSLGKYGKPLARLAVITQA 3772  
QY 389 ---Q-H---N---KE---I--- 394  
Db 3773 LDTKQPIHSGTREARTKNSQDVTVMVVGAHHSIYDAHLVLSMIWRRLYREFIGSQADGILE 3832  
QY 395 ---Y--- 395  
Db 3833 AETSSEGVVPPKSVYKLLRGKDNDESLLFWKEKLRGVSSSQFPASPASWPRVLEHQPST 3892  
QY 396 ---T-I---K---W---S-P---T-G--- 403  
Db 3893 QTLITKVSLPTSSRKLGATVATVAYAAWALTIAHYTADPDVVFQATLSGRETMAQSISH 3952  
QY 404 P---G-T---N---N--- 408  
Db 3953 PESIAGPTIITVPLRIIIDFQTVWSDFLSTLQKDIVRAAYFGQMMGLNSIAHIDNDCRDA 4012  
QY 409 ---P-N---N---AN-L-M-L---A---SAS--- 419  
Db 4013 CGFKSIIVQVQDEGENHVGRAANPFQMSLESIGHFPAPLVEVESESTDVDIRMAYDP 4072  
QY 420 ---F-D---S---T--- 423  
Db 4073 VLVPEKLAHFISDTFTTTTMSNLNLSAANKAKVESIPALSEAHLAELDVTCTPEWILGKAKDE 4132  
QY 424 -VR---L---WD--- 428  
Db 4133 KIRTESHQCLQDLVCRRAQQSPNSQAIDSDGSIYHELDGLSLIAEHLSQLGVRPEAP 4192

```
QY 429 -----V-D----- 430
Db 4193 VCLLFKSKWAVVAMIGIIKAGCGFPLDPSYPHERLEHIISETGSSVIVTSAAYSKLCL 4252
QY 431 -----RGI-C----- 436
Db 4253 SLSVRGIVCDGVSFSTKPLPSTADSPSPFSVRPNQAAVILFTSGTGKPKGVMEHHS 4312
QY 437 ----- 436
Db 4313 VCSALIALGRMGLGQSRVLFQNSYWFVMDLLIDIFGTLVYGCLCIPKEQRMSNLGM 4372
QY 437 -----T-LT----- 441
Db 4373 VQFKVNTMLLSTSVSRMLQPADTPSLETCLTGEAVLSDVDVRWAPKLUHLIAGYPTET 4432
QY 442 -----Q-----E-----P-V-----Y----- 446
Db 4433 CIMSUGELTPSSPANLIGKPVSCQAWINPLKETELAPVGATGELYIQGPTVARGYLHD 4492
QY 447 -----S-----V-----A-----F-----S----- 451
Db 4493 DVLTSAFIVDPQWLTYGTNENQWRRAYKTGDLVFWGQSNLYTVRRKDSQVKIRGQ 4552
QY 452 -----PD-----G-R-----Y----- 456
Db 4553 RVELAEIEEVIROHIPDVTVCVDLLSSDDQNTRIILGAVLIGDRALGDPEDLEVIGYM 4612
QY 457 -----L-AS-----GS----- 461
Db 4613 DLLKSHIIPALEASLPHHMIPEAVVPVQLPTLGSGLKDRKTVRRVAGPLAPSLPOASAR 4672
QY 462 ----- 461
Db 4673 HPNQPTVTHQKLLRLQWCKILPQDESAYVKODNFIIGGDSIAAKLVALLRQHGISL 4732
QY 462 -----F-----D-----K----- 464
Db 4733 AVAEIFTRPLEAMSSLIDEHNFVSHAGILSDVTRTSGVMRQTNLIAGRHSMAVEKS 4792
QY 465 -----C-----VH-----I-----W-N----- 470
Db 4793 RECDNSTLPCTEQOMELAGTEAFTGAHSAQFIFRLPEKIDLRLOAAFDHCADWYPNLR 4852
QY 471 TQ-----V-CLHY--LN-----G-QV----- 482
Db 4853 TQIHKDADTGRLLHDISPVGKVPWSC-HYSDDLNTVLSHDKPPPLGDLPLHRVTIMRH 4911
QY 483 -----L--LN-----LG----- 487
Db 4912 RDPTEMLVTLNHAAYDAWSLRMLLEHYTEAVANPDYEPSYSLGWTAFVLHTENTKEAS 4971
QY 488 RS-----I----- 490
Db 4972 RSFWSSYLDVKPARLMFNVLNPNRQDRLEARIINPKRVLSQATAATVLLAGLTLV 5031
QY 491 -----C-----LYT----- 494
Db 5032 ARVCDTRDVLHALLTGRTLPLAGIENCPGPTITKVPRLPLMDQDLVLTLELDSVAKKIT 5091
QY 495 -----L--PH-H--L-----V-----VI-P----- 503
Db 5092 AELMRVPHPSHSGLSAIREFIPQAEQTTSSGKFHAGSLGRPLDLVIFPKGGLDLLGK 5151
QY 504 -----LVA-----L----- 510
Db 5152 HGLGLQNEGFRLVAPPSGGLSMECALVDDDDKRSITISVDVSLVMDQRAATOBDVIELV 5211
QY 511 --L--VL-K 514
Db 5212 HSLQGIPTK 5220
```

```
RESULT 10
S58870
reelin precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Dec-1996 #sequence revision 06-Dec-1996 #text_change 21-Jul-2003
C:Accession: S58870; S71844; I49297
R:D'Arcangelo, G.; Miao, G.G.; Chen, S.C.; Soares, H.D.; Morgan, J.I.; Curran, T.
Nature 374, 719-723, 1995
A:Title: A protein related to extracellular matrix proteins deleted in the mouse mutant
A:Reference number: I49297; MUID:95231649; PMID:7715726
A:Accession: S58870
A:Status: nucleic acid sequence not shown
A:Molecule type: mRNA
A:Residues: 1-3461 <DAR>
A:Cross-references: EMBL:U24703; NID:G902486; PID:G902487
R:D'Arcangelo, G.
A:Reference number: S71844
A:Accession: S71844
A:Molecule type: mRNA
A:Residues: 1-215, 'T', 217-1905, 'S', 1907-3355, 'V', 3357-3391, 'N', 3393-3461 <DA2>
A:Cross-references: EMBL:U24703; NID:G902486; PID:G902487
F:1-27/Domain: signal sequence #status predicted <SIG>
F:28-3461/Product: reelin #status predicted <MAI>
F:1769-1795/Domain: EGF homology <EGF>

Query Match 70.4%; Score 2568.3; DB 2; Length 3461;
Best Local Similarity 11.5%; Pred. No. 7,7e-53;
Matches 395; Conservative 94; Mismatches 13; Indels 2947; Gaps 362;

QY 1 M-----SI-----S-----S-----D-----EV----- 8
Db 1 MERGCWAPRALVLAIVLLATLARAATGYPRFPFPLCTHGELEGGQGEVLISL 60
QY 9 ---N-----F---LV-YRY---LQ---ES-----GF---S-H 23
Db 61 HIAGNPTYVYVQGEYHVTISTTFDGLLVTLG-YTSTSIQSSQSGSSAFGIMSDH 119
QY 24 ---S-A-----F-----T-----F-----G-I-K----- 31
Db 120 QFGNQPMCSVAVSHVSHLPTNLSPFWIAPPAGTGCNVFMATATHRGQVIFKDALAQQLC 179
QY 32 -----SH-----I-----S-Q-----SNI-----N-----GAL----- 43
Db 180 EQGAPTEATAYSHLAIEHSDSVILRDDFDSYQQLLELNPINWECNCEMGEQGTIMHGN 239
QY 44 -V---P-P-P-----A-AL-I-----SI-----IQ--- 54
Db 240 AVTFCEPYGRELTTCLNTTASVLQFSIGSGSCRFSDPSITVSYAKNNTADWILE 299
QY 55 -----KG---LQY---V-BA-E-----VSIN----- 67
Db 300 KIRAPSNVSTVTHILYLPPEAKGESVQFQWKQDSLVRGVEVEACWALDNLV-INSARE 358
QY 68 ---ED-----GT-LF-----DG-----R----- 76
Db 359 VVLEDNLDPDVTGNWLFPPGATVKHSCQSDGNSIVFHGNEGSEFPATTRDVLSTEDIQ 418
QY 77 -----P---I-----ES-LSLI-----DA----- 86
Db 419 EQWSEEFESQPTGWDILGAVWGADCGTVSGSLVFLKDKGERKLCITPYMDTGYGNLRFY 478
QY 87 -VM-----P-----D-----V-----V----- 94
Db 479 FVNGGICDQGVSHENDIILYAKIEGRKEHIALDTLTYSSVKVPSLVSVVINPELQTPATK 538
QY 95 ---RQ---QAY-R---D-----K---LAQ-----Q----- 106
Db 539 FCLRQKSHQGYNNRNWAVDFHVLVPLPSTMSHMIQFSINLGCCTHQPQGNVSLFSTNH 598
QY 107 -QA-----A-----AAA-----AA----- 114
Db 599 GRWSLLHTECLPEICAGPHLPSTVSYSENYSWNRITPLPNAALTRDTRIRWQTGP 658
```



115 QY A AASQ O GS 123  
659 DB ILGNMAIDNVYIGPSCLKFCGRGQCTRHGCKCDPFGSPACEMASQTFPMFISBFGS 718  
124 QY AK N G E N 129  
719 DB ARLSSVHNYSIRGAESVFCGVLASGKALVFNKGRQLITSLDSSQSRLQFTLRIG 778  
130 QY T A NG E E N G G 137  
779 DB SKSVLSTCRAPDOPGEGVLLHYSYNGITWKLLEHYSYNYHPRILISVELPDDARQFGI 838  
138 QY A H T I AN N H 145  
839 DB QPRWQYHSSQGEDVWAIDEIVMTSVLNSISLDFNLVEVTQSLGLGNVQPYCGHD 898  
146 QY T D M MEVDG D V E 156  
899 DB WTLCTFGDKLASSMRMYVETQSMQI GASYMIQFSLVMGCGQKYTPHMDNQVKLEYSANH 957  
157 QY I P PN K A VVL R 166  
958 DB GLTWHLVQEECLPSPSCQEFBSASIIYHASEFTQWRVTVVLPOKTWSGATFRMSQSY 1017  
167 QY GH E SE V F 173  
1018 DB TAQDEWALDNIVIGQCPNMCSHGSGCDHGVCRCDOGYQCTECPAALPSTIMSDFNP 1077  
174 QY I C A W 177  
1078 DB SSWESDMQBEVIGEVVKEQCGGVSSGSLYFSKAGKRLVSWDLDTSWDFVQFYIQI 1137  
178 QY N P V SD L L ASG 187  
1138 DB GGEACNKPDRBREGILLOYSNNGI QWHLAEMVFSDFSKPRFVYLELPAGKTPCTR 1197  
188 QY SG D S T A R IW 196  
1198 DB FRWKVPVFGEDYDQWAVDDIILSEKQKQVVPVNPPTLPQNFYKPAFDYPMNQMSVWL 1257  
197 QY N LSEN S T S G S T Q L 209  
1258 DB MLANEGMAKNDSCFATTPSAMVFGSDGRFAVTRDLTLKPGVYLQFKLNGICTSQFSST 1317  
210 QY V LR H C IR EG QOD V 222  
1318 DB APVLLQYSHDAGMSWFLLEKGCPPASAAKGCENRELSEPTVYTYGDFEWTITITAI 1376  
223 QY P S N K D 227  
1377 DB PRSLASSKTRFRWIOESSQKNVPPFGLDGVYISEPCPSYCSGHGDCISGVCFCDLGYTA 1436  
228 QY V T S LD W N S E 236  
1437 DB AQCTCVSNTFNHSEMPDRFEGKLSPLMYKITGGQVGTGCTLNDGRSLYFNLGLKREART 1496  
237 QY G TLLAT G SY D G P 248  
1497 DB VPLDTRNISLVQFYIQIGSKTSITTYI TPRARYEGLVVQYNSNDNGILMHLRLDLMFMSF 1555  
249 QY A RIW T K DG 256  
1556 DB LEFQIISIDLPREAKTPATAFR WQPOHQHKSQAOWALGDVLIGVNDSSQTGFQDKLDGS 1614  
257 QY N LAS T L G Q H KG 267  
1615 DB IDLOANWYRIQGVNDIDCL SMDTALI FTENIGNPRYAETWDFHVSSESSFLQWENMGC 1673  
268 QY PIP A L K W NKK N 279  
1674 DB SKP FSGAHI QLOVSLNKGQWLVTBECVPTTIGCVHYTESSTVTSERFQNRWRTVY 1732

280 QY F I LSAGVDKT TI I W DA 295  
1733 DB LPLATNSPRTRFRWIQTNTYTVGAD SWAIDNVILASGCPMCSGRGICDSGRVCVRGFG 1791  
296 QY H T CE 299  
1792 DB GPFCVPVPLPSILKDDFNGNLHDPDLPVGYAERGNLNGETIKSGTCLIFKGEGLRMLI 1851  
300 QY AK Q QF P H SA 309  
1852 DB SRDLCTNTMYVQFSILRFIAKGTPEPERSHSILLOQFSVSGVTWHLMDDEFYFPQTRILFIN 1911  
310 QY P A L D V D W 316  
1912 DB VLPYGAQTNATFRMLQWPYNNKGKKEEIIIDDFTIDGNLNNPVLLLDTDFDPREDNW 1971  
317 QY O S N NT 321  
1972 DB FFYPGNIGLYCPYSKGAPEBSAMV FVSNEVGEHSITTRDLSVNENTIIQFEINVGCS 2031  
322 QY F A S CST 327  
2032 DB TDSSADPVLRFSDRDFGATWHLPLCYHSSSLVSSLCSTEHPSPSTYYAGTTCQMRRE 2091  
328 QY D MC IH VC 334  
2092 DB VVHFGKLHLCGSRFRWYQGFYAGSPVPTWADNVYIGPOCEMCGHGCINGTKCIC 2151  
335 QY K L Q DR P I KT F Q G 347  
2152 DB DPGYSGPTCKISTKNPDFLKDDFEGQLBSDRFLMGGKPSRKCGILSSGNNLFFNEDGL 2211  
348 QY H T N EV 352  
2212 DB RMLVTRDLDLSHARFVQFPMRLCGKGVDPDRSQPVLLQYSLNGLSWSLQBFLEFSNS 2271  
353 QY NA I K W 357  
2272 DB NVGRVIALEMPKARSGSTRLRWOPSENGHFYSPWVIDQILIGNISGNTVLEDDFSTL 2331  
358 QY D TG N LL ASCS D 369  
2332 DB DSRKWLHPGKMPVCGSTGDALVIEKASTRYVVTDDIADVNEFSFLOIDFAASCSTVD 2391  
370 QY DM L KI W 376  
2392 DB SCVAIELEYSVDLGLSWHPLVRDCLPTNVECSRYHLQRLVSDTFNKTITRITLPLSYTR 2451  
377 QY S M KQ DN CV H D 386  
2452 DB SQATFRWHOPAPFDKQQTWADNVYIGDCLDMCSGHGRCVQSGCVDEQMGGLYCDEP 2511  
387 QY L Q H 390  
2512 DB ETSLPTQLKDNFRAPSNQNLTVSGKSLTVCGAVASGLALHFGSGCRLLVTVDLNLT 2571  
391 QY NKE I Y T IKW S PTG P 404  
2572 DB NAEFTQFMYOCLITPSNRNOQVILLEYSVNGITWLLMEIFYDQYKXP GFVNILLPP 2630  
405 QY GT NNP N A N L ML ASA S 419  
2631 DB DAKEIATFRWQ PRHDGLDQNDWADNVILSGSADQRTVMLDTPSSAPVPOHSPAD 2689  
420 QY F D S TV R L 426  
2690 DB AGFVGRIAPFEMFLEDKTSVNENLFDHDDCTVERFCDSPDGVMLCGSHDGRVYAVTHDLT 2749  
427 QY W D V D 430  
2750 DB PTENIMQPKISVCGCKVPEKIAQNIHQVFTDFGVSWSYLVPOCLPADPKCSGVSQPS 2809  
431 QY R G I CI H 436

```
Db 2810 VFPTGKRIITYPLPESLTGNVPRFPYQYSDVQWAINDFYLGCGCLDNCGGHCDCLK 2869
QY 437 -----T-----ITKH-----QE--P-----V-----445
Db 2870 EQICIDPGYSGPNCYLT-HSLKTFKXERFDSBIKPDLMWSLEGGSTCTCGVLAENTAL 2928
QY 446 Y---S-V---A-F-----S--P-----DG-----454
Db 2929 YFGSTVROAITQDLRGAFLQYWGRIGSENNMTSCHRPVCKBGLVDFDSTGGITW 2988
QY 455 -----R-Y-L-A-----A-----SG-----460
Db 2989 TLLHMDFOKYISVRHDYILLPSGALTNTTLRLWQPFVINSGLVSGVGCACAVGTGQHS 3048
QY 461 -----S-----FD-----K-----C-----VHI-W-N-----T-----471
Db 3049 DWSRNHPSQVDTFDEGSSHEENWSFYPNAVRTAGFCGNPSPFLYWPNNKKDKTHNAL 3108
QY 472 -----Q-----V-----C--LH-----Y-----L-----N--479
Db 3109 SSRELIQPGYMQFKIVVGCETATSCDGLHVMLEYTKDARSOWOLVOTQCLPSSNSI 3168
QY 480 G-----Q-----VL--LN-----G-----R-----SI--490
Db 3169 GCSPPQFHEATIYNANSSSKWIKITQLPDHVSSATQFRWIKGBETEKQSWAIDHVI 3228
QY 486 -----L-----L-----G-----R-----SI--490
Db 3229 GEAPKPLCSGHGYCTTGAVCIDCESFGDDCSVFSHELPSYIKDNFESARVTEANWETIQ 3288
QY 491 -----C-----LY-----T-----T-----494
Db 3289 GGIVGGCGLAPYAHGDSLYFNGCQIROAATKPLDLTRASKIMFVLQIGSPAQTSCNS 3348
QY 495 -L--PH-----HLV-----V-----IPL-----V-----505
Db 3349 DLGSPHTADKAVLLQYSVNGITWTHVIAHQHPKDFQAQVRSVHVPLEARMKGVLLRWQ 3408
QY 506 -----AL--IB-LIV-----LK 514
Db 3409 PRHNGTGHQDQWALDHVEVVLVSTRKQNYM 3437

RESULT 11
T12117
polyprotein - fava bean dsRNA replicon
C:Species: Vicia faba (fava bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 09-Jul-2004
C:Accession: T12117
R:Pfeiffer, P.
J. Gen. Virol. 79, 2349-2358, 1998
A:Title: Nucleotide sequence, genetic organization and expression strategy of the double
A:Reference number: Z17424; MUID:98451319; PMID:9780039
A:Accession: T12117
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-5825 <PFE>
A:Cross-references: UNIPROT:O82731; EMBL:AJ000929; NID:g3184155; PIDN:CAA04392.1; PID:g3
A:Experimental source: virion; cultivar 447
C:Comment: This gene product may be cleaved into several proteins including helicase and
C:Genetic:
A:Genome: dsRNA replicon
C:Superfamily: fava bean dsRNA replicon polyprotein
Query Match 70.3%; Score 2566.1; DB 2; Length 5825;
Best Local Similarity 9.3%; Pred. No. 2.6e-52;
Matches 441; Conservative 57; Mismatches 13; Indels 4239; Gaps 379;
QY 1 M-----SIS-----S-----D-----E-----V-----N--9
Db 949 MATRPSITQAPVWKSNTLLIDMNLIRMDPMWLITECLRRKITTAYVTPMELDONPD 1008
```

```
QY 10 -----F-----LV-----Y-----13
Db 1009 QTILDEVOFRKVENQIYVITAWSDRPIRVTTENYNNWTKVDAFVIGENTYPIDTVREAN 1068
QY 14 -----R-----Y-----L-QE--18
Db 1069 GISLRIDLAASSTPMSSEWVETVGDSEISVPTVYVYKPGKRKLSMGSEKKVKLHREL 1128
QY 19 -----SG-----PS-----HS-AF-----26
Db 1129 YRMMIRNVSGTNSFKELVDYIGISNMKNYNIMVMVVSFSDISDDTIRHTAAAFVSKMIT 1188
QY 27 -----T--F-----28
Db 1189 RANSMQLTERTVLESVNTARGFLASLLISOAETMLKITGDETMLTITVNRWMSKSEWVRL 1248
QY 29 -----G-----G-----IK-----31
Db 1249 KGMLMNLISLDIDNLMCYSRPVIHRLRAQETLELEPTGTCPHKTPDFVIKECSMCECCG 1308
QY 32 -----S-----H-----33
Db 1309 VNPAGIAGFCSCSLPTDKHQCDHPCKHAHESEAGTRCSCLLPIAGEACPCGVNRQIE 1368
QY 34 -I-----S--O-----SNIN--40
Db 1369 SEILPENSESADAEEQNRVKNKXPRPDREKGNRNNRNDSSRRTDANHHISNVYHH 1428
QY 41 -----GALV--PP--AAL-----I-----S-----I 52
Db 1429 GHNKPORQGA-TQPPKRSALHPENDNDTDPDSTPVLITADPTTNPGERQAPGCDNI 1487
QY 53 -I-----Q-----KG-----L--O-----Y-----V 60
Db 1488 PHENDIPGSPSTTQSSPPDNTYSGPEHNRIQNHRIIAIQTNEFFGDLAYAGNDIAV 1547
QY 61 -----E-----A--EV--S-----I--N 67
Db 1548 KKNSHKLPNPENQRLRSCFCQFMNMGDSIYATENLEVIKTYNTSGEGGYCGYNALKILYPN 1607
QY 68 -----E--D-----69
Db 1608 LDLTLEMQEIVGSETQFSDWEIMRVAQAKQLNLIIVTERCALVNSKSVCSNEFGVICHCR 1667
QY 70 -G-----TL-----F-----D--G-----R-----P--77
Db 1668 HRGVMLHWEAALAIQKGFADYHPTTNALTREDLLDFAKSSGLNKNKVRSFVLGDPRL 1727
QY 78 -I-----ESLS-LI-D--A--VM-----P-----D-----90
Db 1728 KLQTEMHESLSKVVEDVNPAGFVKITKGSTHYVTVNDRNRLGYSPQYGTFTAQITDNOEL 1787
QY 91 -----VVQT-----R-----Q-----QA-----98
Db 1788 VELLILAYSAPARVIHTDFYDRPESVPADATQGIHDYKQLVRDIAEVNQACAGDLIKS 1847
QY 99 -Y--R-D-K-----L-AO-----Q-----O--107
Db 1848 ELIQNRVDCKEYINNHLFPVKORTKLPKGLISGLKVLGIQSTYSYVDVSHVIGVIRNQL 1907
QY 108 -----A-A-----A-----A-A--A--A--114
Db 1908 WCTCVTEKCAKATKIKVDLYRNKNTGSLRALFGIPRWDPRDRIKTLLEKATAVDAIAGWG 1967
QY 115 -----A--AAAS-----QQG-----S-----123
Db 1968 KSTEIVKLVNQDCTVVAQTSAAVSNILEKQGGKKNMCKVMSIEKMTQVQNTPTLVLD 2027
QY 124 -A-----KN-----G-----E-----128
Db 2028 EASMITWETLSLITGPOVENVYLGNTLOICVLDWYRTGSRATKSIIEQAGIIRHYTT 2087
QY 129 ---N-----T-A--N-----G--EE-----135
```

Db	2088	HRIGNPLARELSLVTKELITNAKHETNCTKSWDAVRWAELOTSIAGSLSEBPVILCFYNA	2147	Db	3167	YGRSYLMTQVYTERYVNVVINKNGLDLEAVCIAALTMKHIRNEQLEHTMGLMETFRYVDT	3226
Qy	136	-----N-----G-A-----H-----T-----	140	Qy	266	-----K-G-P-----I-----F-----	270
Db	2148	VRAVMYLVKVCGRVDTHHKFOGLEADNVVLQWCPTGPTGPRITLDKHQCLSAATRAKON	2207	Db	3227	DSITDIVKSKNLEKVMSTANTMOEKLGPASFTGLLDEANTQORVVVFIDERTNQSR	3286
Qy	141	-----I-----ANN-----HT-----DM-----M-----	149	Qy	271	-----A-----LKW-N-----K-----	277
Db	2208	LWISVNEYSNVPLHKRMGATIGGSKHTQPDENNADKTLQVSHLLSKVMIYKDKT	2267	Db	3287	VMSWDSNIKMSDESGIKKMWASTADNLDNMNIGMIQEVYKVLCKREDIPKTKKACPGF	3346
Qy	150	-----E-----VD-----G-----	153	Qy	278	-----GN-----	279
Db	2268	IERLECSRLQSOBEGNLVMLKNLTLHLDVDYVSGNDQHSRLRQSGKLAKEFKKMSLSLFS	2327	Db	3347	KQVVDCAKAVAOQKEARKTDPGARVLVLEPIPLIGNVDITRAATOGTMVTKNENVIRSE	3406
Qy	154	-----D-V-----EI-----	157	Qy	280	-----F-----L-S-----AGVD-K-----T-TII-----	292
Db	2328	KEPAPDHVSSSTLGLKAEIVDQANANKDRINSQIMNIYTVQKNLAINSGSGIAAATPVVL	2387	Db	3407	ITFRARQSRRAQTSRLRSYLDNWNEMVKAYSALEAG-DLKVMVRELSRIALIKSTSTP	3465
Qy	158	-----P-----N-K-----AVV-----LR-----	166	Qy	293	-----D-----A-H-----T-----GEA-----KQ-----Q-----F-----	304
Db	2388	TAWLAQKPGITKAECSEDFPTETLTANILMLGVNVLKIQLDFENDKVILGSGGLRSFVAD	2447	Db	3466	TWHLKEDLEAMLDVPNLTRAHLEITKLIGDAVVSWEQSKAEKKQNEVDFIKSLDLGLST	3525
Qy	167	-----GH-----E-S-----E-----V-F-IC-----	175	Qy	305	-----P-----	305
Db	2448	SSEIKOLYNQKFNCCGHYVLKYRNFDTESVKKMNFERVLWTVNYFGVLESGLHSIW	2507	Db	3526	LNDHDLNIVLSLLAVDRPSKVKKDLLTNTLGSQMSVNEAVMLRRKIRQOWCESASNGQL	3585
Qy	176	-----AW-----N-----P-----V-----SD-----LLA	185	Qy	306	-----PH-----S-----A-----	309
Db	2508	TIENELYKMQSGYGCSCGGMVIFHNKPIIVGDPSYNTGNSRTVFNPSKQSDKCPILA	2567	Db	3586	VDHHDHANSVSTEQNOGKLNILSLHEBAGICDINLSDASPLUKIRASSYLPBAITTCVQ	3645
Qy	186	-----SGS-----GDS-TA-----RI-----W-----	196	Qy	310	-----P-----ALDVD-----W-----OS-----	318
Db	2568	YLRWNVDSTGVKLPYDPVGRGDFMEGDNVMAALVIERVLKGMMAAKLKNMLTW	2627	Db	3646	TPGMVELTSPKVEIIEAMDNDCEWRCIEKYVTQNIIEPHFRITNGLRALMQOSKMLTES	3705
Qy	197	-----N-L-S-----E-----NS-----T-----SG-S-----TOLV-----	210	Qy	319	-----N-----N-----T-----F-----AS-----C-----STD-----	328
Db	2628	TNODLCNFYDNNSPFLFDSLIKARLESNGFTNFNTLVNFBVGTGRLEFNGFTNFNTLVN	2687	Db	3706	QAIIVCOLLGLACCLIQNGETGTVYNFAPNKPQVQLMRLSQDASDHCVLINLGLADGVK	3765
Qy	211	-----L-BH-C-----I-----	215	Qy	329	-----M-----C-----I-----HV-----C-----	334
Db	2688	VNSMDLRCPILCTLDKRYLAALMDQDNLIWTVQISGKPIHTPAGVTRVSNKAFKTMAYR	2747	Db	3766	RUSPENIAKENMEQLEHVCVDNKPIVAVGENPYACVSHVHDNEDLARLITSCSQPYEKI	3825
Qy	216	-----R-E-----G-----Q-----D-----	221	Qy	335	-----K-----	335
Db	2748	QMLNRLEFKLTWQLSRDKPMGAPNPRGPIILKPMQAHEADNKVAVAARKYKQTOAAMM	2807	Db	3826	PVLDCGRFEMLPVNNVMARMTRKTSSATLLKKVTPITTGAGSLALTRNYKATPMVTTKPG	3885
Qy	222	-----V-----P-SNK-----D-V-----T	229	Qy	336	-----L-----G-Q-----	338
Db	2808	KLKSRMYLTAWVNGSEYASIMQOHNVITTSYNNPENMLGLVDLYVAQAIHNTGYT	2867	Db	3886	DVLVLRDHCEKKVMCTHTTIVDTGGVQIILEITDDIGDGVCLDLGLHCLTPADRTPR	3945
Qy	230	S-----L-----DW-----N-----SE-----	236	Qy	339	-----DRP-----IKT-F-----OG-----H-----T	349
Db	2868	SALYITNACTAVLCHGHWDFSCPPDDGWSRFLSSTHTDTITNLADIKSIILDSIAKLD	2927	Db	3946	RDRPNTSNEVWSLNIETORYLEATKGLSHVQGGEGKLHIYHFDNREHHMYDEHYDA	4005
Qy	237	-----GT-----	238	Qy	350	-----N-E-----VN-----A-----IK-W-----D-P-----	359
Db	2928	KQBELSEETAVLKEKERSERIHNQILRTSDSWYESYQNVKRGIVIVSVNSCGISTES	2987	Db	4006	MLDENIVIELDPTVQVPTGVNIVNLICIRRMFLAMLYGIPTVWVAYCDDNPSAAQLAWFL	4065
Qy	239	-----L-----A-----TGS-----Y-----	245	Qy	360	-----	359
Db	2988	IEKIMETTGAEFLMVCVPTLTPSDNTTAHRVGTGNSNNKITYPGARNTLTNPETVYCM	3047	Db	4066	KYYKMGHKVFNNNVTTEDLTNTRAMEIINALFSTYQHTTTTNTTTRVHLRNKAETILSKBK	4125
Qy	246	-----DGF-----A-----R-----	250	Qy	360	-----T-----G-----N-----DD-----L-----L-----	364
Db	3048	KSGKRATVOSQTKD-FDMISLHTTTVAMIASVQLKVTITTRFNEHAMPKALPIPFSPFS	3106	Db	4126	VSEELFNTTIDISVFTAVLMGLENEVLKSNQGTCTVFDLIPETPEPPLOQDMLESMLV	4185
Qy	251	-----I-W-----T-KDG-----NL-----A	259	Qy	365	-----A-----S-----CS-----	370
Db	3107	HOSDEKVVQVPRFTCNFGDGSNSLIITKRRKLVSKALLAKMNERALRHDSNLKDLOA	3166	Db	4186	DIKQCAVHVPTGTKVLRVVSLSHEIACSTKYKGVDPKDDDEMAGASNLTSQSPASDK	4245
Qy	260	S-----T-L-----G-----Q-H-----	265	Qy	371	-----M-----T-----L-----KI	375
				Db	4246	ADPGSSQLSSCBENMSQTGTQPVACVHIBPCDHAANASNDVLKEITQPEQNNHDIISKI	4305

```
QY 376 -----W-----S-----M----- 378
Db 4306 LATAQAKSNPEKSKGATSDLAQIQLCLDSDPGMGSDQADQAKKLEVMNFSDEVT 4365
QY 379 -----K-Q-----D----- 381
Db 4366 LSGIHRQRYNSQGYLQVDIRSWAITYKIGEGHSDVPTNNPGWMLAPPPMRKLG 4425
QY 382 -----NC-----V-H-----D-L-----Q----- 388
Db 4426 VYKEHEIKVTMYAGNCAYRLNRIKREPKSVFTINVKPEVMDLILTYTSANNKQYPTN 4485
QY 389 -Q-----H-----N-----K-E-I-----Y----- 395
Db 4486 LQRLADKYLHGRVSHVYVYEGYNRDLQTCVELTKRMERVIIVVTDAVDVYNQLDADIQV 4545
QY 396 -----TIK-----W-----SP-----TG-----P----- 404
Db 4546 REKIITKVTQSPEPYRAIWRFMILPANIDPECELYILNGREMAPNDMELVDVRMEESF 4605
QY 405 -----G-----T-N-----N----- 408
Db 4606 EDFDIVSALHCOAFNCCGVAKYNHAKFTDYNTFNDIYEAGNSYGGDELYMSRYKVXW 4665
QY 409 -----P-NA-----NLML-----A-----SASF----- 420
Db 4666 WLYYILPGGAWYGIEGKWLNYALDITICLEKEGERTRIDHVKGLFCSANFHVADPTKG 4725
QY 421 -D-STV-----R-----L----- 426
Db 4726 IYDSTVDSVDGIPHOYAMQINGHSVTENPVRHSTGESTGNPRINSDGEPWQLNPN 4785
QY 427 -W-D-----V-D-----R-G-I-----C----- 434
Db 4786 NWLLDPLKSKEMKIQKFLDSSVKIKESQSLPIQLGPLKPIVVFVNAQACEAIGVSM 4845
QY 435 -----I-H-TL-----T----- 439
Db 4846 NNTGTLVVEVNIQAOKLAARNEPLOHQTLIKSHLTLMTSTGRHASKWPEYTHLMKADLQ 4905
QY 440 -----K-----H-Q-----E-----P-----V----- 445
Db 4906 NCQYVEGSATSLQWKRETSQAQVYVTEERDVFRIPQALDHSVGLTKLVKVTGNE 4965
QY 446 --YS-----V-----A-----F-----S-----P----- 452
Db 4966 LGLSSSGVLARVLKLNDRLANHWHLETOHLQKLEFPDRRIISHEESSGMTKPRIGOT 5025
QY 453 -D-----D-----G-----R----- 455
Db 5026 LYLADVLTAAGIBCVNHTTTSREWKEALKNPLTKVGIMPLTNSVLVSGKLCVQPTIPR 5085
QY 456 -Y-----Y-----L-----AS----- 459
Db 5086 CYAMLCQGVKLIKFPETRGGTGPSNVANVGSNPTTAGPGDTPPLITSWNKWSGAS 5145
QY 460 -----G-S-----FD-----K----- 464
Db 5146 NQSPFGVSAYQDKLREKKIETPDRLRLVLEMTLFDAYALMSHGLTLDRKLDNGGNWH 5205
QY 465 C-----V-----H-----T-----W-----N----- 470
Db 5206 CATDATWKSDFEIEPVGGDLLPSLNTSALRHNETNARVIDLWDDTDLRDLWLTLYAPKNP 5265
QY 471 -T-----Q-V-----C----- 474
Db 5266 MKLTSRVSPPGGKINLKLTLLNRCQTRPVPQVMGGENAVTGRGLSVLPLRREPNNVT 5325
QY 475 --LH-----Y----- 477
Db 5326 HELHKFTAYYRGWERVLKDFKANTITISDADVKTWLSRRSDWKALATSTIKMLTGLP 5385
```

```
QY 478 -----L-N----- 479
Db 5386 SNPMNAVNVHTESLLKANPIWYRQTOGRIIWNQPKELCALMSPAFIAIKRLEKVL 5445
QY 480 -----G-----QV-----L-----L----- 484
Db 5446 DEIVYTDGLTPDMLSARARTIQDYVVFEDDLVIQDRQTOELIDLEFQVMLDLGLDINLA 5505
QY 485 NL-----G-----R-----SI-----C----- 491
Db 5506 NLWRLVHNKWRPKQGSWGLDAMBLTQOATLTALGNATNLCSHVSFVIEHROAIKLMFV 5565
QY 492 -----L-----Y-----T-----L-----P----- 496
Db 5566 LGDDNITFMSAPNLTYYKRLMSERYNMRSKQVSRNVCTFCSLLCYRNSFGHCEVGPDP 5625
QY 497 -----H-----H-L-VV-I-P-----L-V-----A-L-----IE-L----- 510
Db 5626 VRLRHRFEVTGGNHVELPVAVKMHPHNVIRIGKTFNGRYLHVMEENEAEAEFVWELRNH 5685
QY 511 -----LVLK 514
Db 5686 ANRSYILVLK 5695
```

## RESULT 12

## S02392

alpha-2-macroglobulin receptor precursor - human

N;Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor

C;Species: Homo sapiens (man)

C;Date: 14-Aug-1998 #sequence revision 14-Aug-1998 #text\_change 09-Jul-2004

C;Accession: S02392; S30027; I37998; A39210; S12538

R;Herz, J.; Hamann, U.; Rogne, S.; Myklebost, O.; Gausepohl, H.; Stanley, K.K.

EMBO J. 7, 4119-4127, 1988

A;Title: Surface location and high affinity for calcium of a 500-kd liver membrane protein

A;Reference number: S02392; MUID:89210795; PMID:3266596

A;Accession: S02392

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-454 &lt;HR&gt;

A;Cross-references: UNIPROT:Q07954; EMBL:X13916; NID:G34338; PIDN:CAA32112.1; PID:G34339

R;Kristensen, T.

submitted to the EMBL Data Library, October 1990

A;Reference number: S30027

A;Accession: S30027

A;Molecule type: mRNA

A;Residues: 3275-3864 &lt;KRI&gt;

A;Cross-references: EMBL:X55077

R;Herz, J.; Kowal, R.C.; Goldstein, J.L.; Brown, M.S.

EMBO J. 9, 1769-1776, 1990

A;Title: Proteolytic processing of the 600 kd low density lipoprotein receptor-related p

A;Reference number: S12538; MUID:90269210; PMID:2112085

A;Contents: annotation; site of proteolytic cleavage

R;Kutt, H.; Herz, J.; Stanley, K.K.

Biochim. Biophys. Acta 1009, 229-236, 1989

A;Title: Structure of the low-density lipoprotein receptor-related protein (LRP) promoter

A;Reference number: I37998; MUID:90089395; PMID:2597675

A;Accession: I37998

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-11 &lt;RES&gt;

R;Cross-references: EMBL:X15424; NID:G34408; PIDN:CAA33464.1; PID:G34409

R;Strickland, D.K.; Ashcom, J.D.; Williams, S.; Burgess, W.H.; Migliorini, M.; Argraves,

J. Biol. Chem. 265, 17401-17404, 1990

A;Title: Sequence identity between the alpha2-macroglobulin receptor and low density lip

A;Reference number: A39210; MUID:91009181; PMID:1698775

A;Accession: A39210

A;Status: preliminary

A;Molecule type: protein

A;Residues: 150-166;234-238, 'X', 240-245, 'X', 247-252; 'G', 686-695; 902-916; 1096-1109; 'S', 17

C;Genetics:

A;Gene: GDB:LRP1; APR; LRP; A2MR

A;Cross-references: GDB:119694; OMIM:107770

A:Map position: 12q13.1-12q13.3  
 C:Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated protein (see PIR:A39875).  
 C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding site; beta-hydroxyaspartic acid; calcium binding; glycoprotein.  
 C:Keywords: beta-hydroxyasparagine; signal sequence #status predicted <515K>  
 F:1-13/Domain: signal sequence #status predicted <515K>  
 F:20-3943/Product: alpha-2-macroglobulin receptor 515K chain #status predicted <515K>  
 F:27-64/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
 F:72-108/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
 F:115-148/Domain: LDL receptor ligand-binding repeat homology <EG1>  
 F:154-188/Domain: EGF homology <EG2>  
 F:198-239/Domain: LDL receptor YWTD-containing repeat homology <YW01>  
 F:240-281/Domain: LDL receptor YWTD-containing repeat homology <YW02>  
 F:292-334/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
 F:335-378/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
 F:379-420/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
 F:421-468/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
 F:478-519/Domain: EGF homology <EG3>  
 F:571-613/Domain: LDL receptor YWTD-containing repeat homology <YW07>  
 F:614-659/Domain: LDL receptor YWTD-containing repeat homology <YW08>  
 F:660-710/Domain: LDL receptor YWTD-containing repeat homology <YW09>  
 F:711-752/Domain: LDL receptor YWTD-containing repeat homology <YW10>  
 F:753-799/Domain: LDL receptor YWTD-containing repeat homology <YW11>  
 F:807-842/Domain: EGF homology <EG4>  
 F:854-890/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
 F:895-931/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
 F:936-971/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
 F:976-1011/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
 F:1015-1051/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
 F:1062-1097/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
 F:1104-1140/Domain: LDL receptor ligand-binding repeat homology <LDL9>  
 F:1145-1182/Domain: LDL receptor ligand-binding repeat homology <LDLA>  
 F:1185-1221/Domain: EGF homology <EG5>  
 F:1227-1261/Domain: EGF homology <EG6>  
 F:1269-1308/Domain: LDL receptor YWTD-containing repeat homology <YW12>  
 F:1309-1355/Domain: LDL receptor YWTD-containing repeat homology <YW13>  
 F:1356-1398/Domain: LDL receptor YWTD-containing repeat homology <YW14>  
 F:1399-1445/Domain: LDL receptor YWTD-containing repeat homology <YW15>  
 F:1446-1488/Domain: LDL receptor YWTD-containing repeat homology <YW16>  
 F:1489-1531/Domain: LDL receptor YWTD-containing repeat homology <YW17>  
 F:1540-1578/Domain: EGF homology <EG7>  
 F:1583-1626/Domain: LDL receptor YWTD-containing repeat homology <YW18>  
 F:1627-1669/Domain: LDL receptor YWTD-containing repeat homology <YW19>  
 F:1670-1713/Domain: LDL receptor YWTD-containing repeat homology <YW20>  
 F:1714-1753/Domain: LDL receptor YWTD-containing repeat homology <YW21>  
 F:1754-1796/Domain: LDL receptor YWTD-containing repeat homology <YW22>  
 F:1797-1846/Domain: LDL receptor YWTD-containing repeat homology <YW23>  
 F:1850-1886/Domain: EGF homology <EG8>  
 F:1934-1976/Domain: LDL receptor YWTD-containing repeat homology <YW24>  
 F:1977-2019/Domain: LDL receptor YWTD-containing repeat homology <YW25>  
 F:2020-2063/Domain: LDL receptor YWTD-containing repeat homology <YW26>  
 F:2064-2105/Domain: LDL receptor YWTD-containing repeat homology <YW27>  
 F:2106-2151/Domain: LDL receptor YWTD-containing repeat homology <YW28>  
 F:2159-2194/Domain: EGF homology <EG9>  
 F:2199-2241/Domain: LDL receptor YWTD-containing repeat homology <YW29>  
 F:2253-2294/Domain: LDL receptor YWTD-containing repeat homology <YW30>  
 F:2344-2388/Domain: LDL receptor YWTD-containing repeat homology <YW31>  
 F:2389-2429/Domain: LDL receptor YWTD-containing repeat homology <YW32>  
 F:2430-2473/Domain: LDL receptor YWTD-containing repeat homology <YW33>  
 F:2482-2517/Domain: EGF homology <EG10>  
 F:2524-2561/Domain: LDL receptor ligand-binding repeat homology <LDLB>  
 F:2566-2600/Domain: LDL receptor ligand-binding repeat homology <LDLC>  
 F:2605-2639/Domain: LDL receptor ligand-binding repeat homology <LDLD>  
 F:2652-2688/Domain: LDL receptor ligand-binding repeat homology <LDLE>  
 F:2696-2730/Domain: LDL receptor ligand-binding repeat homology <LDLF>  
 F:2734-2769/Domain: LDL receptor ligand-binding repeat homology <LDLG>  
 F:2774-2812/Domain: LDL receptor ligand-binding repeat homology <LDLH>  
 F:2816-2853/Domain: LDL receptor ligand-binding repeat homology <LDLI>  
 F:2858-2897/Domain: LDL receptor ligand-binding repeat homology <LDLJ>  
 F:2904-2939/Domain: LDL receptor ligand-binding repeat homology <LDLK>  
 F:2944-2980/Domain: EGF homology <EG11>  
 F:2986-3021/Domain: EGF homology <EG12>  
 F:3029-3068/Domain: LDL receptor YWTD-containing repeat homology <YW34>

F:3069-3113/Domain: LDL receptor YWTD-containing repeat homology <YW35>  
 F:3114-3156/Domain: LDL receptor YWTD-containing repeat homology <YW36>  
 F:3157-3200/Domain: LDL receptor YWTD-containing repeat homology <YW37>  
 F:3201-3241/Domain: LDL receptor YWTD-containing repeat homology <YW38>  
 F:3242-3284/Domain: LDL receptor YWTD-containing repeat homology <YW39>  
 F:3294-3330/Domain: EGF homology <EG13>  
 F:3334-3369/Domain: LDL receptor ligand-binding repeat homology <LDLL>  
 F:3374-3408/Domain: LDL receptor ligand-binding repeat homology <LDLM>  
 F:3413-3448/Domain: LDL receptor ligand-binding repeat homology <LDLN>  
 F:3453-3489/Domain: LDL receptor ligand-binding repeat homology <LDLO>  
 F:3494-3531/Domain: LDL receptor ligand-binding repeat homology <LDLP>  
 F:3536-3570/Domain: LDL receptor ligand-binding repeat homology <LDLQ>  
 F:3575-3609/Domain: LDL receptor ligand-binding repeat homology <LDLR>  
 F:3613-3647/Domain: LDL receptor ligand-binding repeat homology <LDLS>  
 F:3654-3690/Domain: LDL receptor ligand-binding repeat homology <LDLT>  
 F:3695-3731/Domain: LDL receptor ligand-binding repeat homology <LDLU>  
 F:3741-3776/Domain: LDL receptor ligand-binding repeat homology <LDLV>  
 F:3785-3822/Domain: EGF homology <EG14>  
 F:3828-3860/Domain: EGF homology <EG15>  
 F:3868-3911/Domain: LDL receptor YWTD-containing repeat homology <YW40>  
 F:3912-3969/Domain: LDL receptor YWTD-containing repeat homology <YW41>  
 F:3944-4044/Product: alpha-2-macroglobulin receptor 85K chain #status predicted <85K>  
 F:3944-4044/Domain: 85K chain extracellular #status predicted <EXT>  
 F:3970-4012/Domain: LDL receptor YWTD-containing repeat homology <YW42>  
 F:4013-4056/Domain: LDL receptor YWTD-containing repeat homology <YW43>  
 F:4057-4099/Domain: LDL receptor YWTD-containing repeat homology <YW44>  
 F:4100-4142/Domain: LDL receptor YWTD-containing repeat homology <YW45>  
 F:4151-4182/Domain: EGF homology <EG16>  
 F:4200-4231/Domain: EGF homology <EG17>  
 F:4236-4267/Domain: EGF homology <EG18>  
 F:4272-4303/Domain: EGF homology <EG19>  
 F:4308-4339/Domain: EGF homology <EG20>  
 F:4344-4374/Domain: EGF homology <EG21>  
 F:4377-4408/Domain: EGF homology <EG22>  
 F:4421-4444/Domain: transmembrane #status predicted <TM>  
 F:4445-4544/Domain: intracellular #status predicted <INT>  
 F:166,2998/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
 F:2958/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted  
 F:4075,4125,4276/Binding site: carboxylate (Asn) (covalent) #status predicted

Query Match 70.2%; Score 2562.6; DB 1; Length 4544;  
 Best Local Similarity 10.4%; Pred. No. 1.9e-52;  
 Matches 414; Conservative 79; Mismatches 16; Indels 3474; Gaps 366;

QY	1	MS-----I-----SSDE-----V-----N--F-L-----11
DB	88	MSRLCNGVQDMGSDGPHCRELQNC SRLCGQHCVPTLGGPTCYCNSSFQLQADGKT 147
QY	12	-----VY-----R-----Y-LQ-----E-----18
DB	148	CKDFDECSVYGTCSQLCTNTDGSFICGCVGYLLQPDNRSCAKQNEPVDPRPPVILLANSQ 207
QY	19	-----SG-----F5-----H--SA-----P 26
DB	208	NILATYLSGAQVSTITPTSTROTAMDFSYANETVCVHVHGDSSAQTOLKCARMPGLKGF 267
QY	27	-----TFGK-S--HLSQ-----S-N-----I-----NG-----41
DB	268	VDEHTINISLSLHVHVEQMAIDMLTGNFYFVDDIDRRIFVCNRRNGDTCVTLLELLEYNPKG 327
QY	42	AL-----VP-----P-----A-A-48
DB	328	IALDPAMGKVFPTDYGQPKVBERCDMDGQNRKLVDSKIVFFHGITLDLVSRLVYWADAY 387
QY	49	--LI-----SIQKG-L-Q-Y--V-E-----A--EVS-I-----N 67
DB	388	LDYIEWVDYEGKGRQTIQ-GILIEHLYGLTFVFNLYATNSDNANAQOKTSVIRNREN 446
QY	68	--E-----D--CTL-----72
DB	447	STEQVTVTRVDKGGALHYHQRQRPRVRSHACENDQYKPGGCSDICLLANSHKARTCR 506
QY	73	---F---DG-----R-----PIE-----SL--81

Db	507	RSFSLGSDGSKCKKPEHEFLVYKGRPGIIRGMDMGAKVDPDEHMIPIENLMPALDF	566	Db	1586	LYARQMEIRGVLDAPYYNYIISPTVPIDNVTVLDYDAREQVYVSDVTRTOAKRAF	1645
QY	82	-----S-LI-----D-----AV-M-----PD--	90	QY	218	G-G-----QVPSNK-----D-V-----TS-----LD-----	232
Db	567	HAETGIYFADTYSYLIHQKIDGTEREILKDGHNVEGVAVDMGMDNLYWTDGPKKT	626	Db	1646	GTGVETVVSADLP-NAHGLAVDWVRNLFWTSYDTNKKQINVARLDGSKFNAVOGLRQP	1704
QY	91	V-V-----QTR-----	97	QY	233	-----W-----N-S-----EG--TLL--A--TG-----S-Y-----	245
Db	627	ISVARLEKAAQTRKTLIEGKMTHPRAIVVDPLNGMYWTDWBEEDPKDSRRGRLERAMDG	686	Db	1705	HCLVVPURGLKYWTGDGNI SWANMDGSRNLTLLFSQKGPVGLAIDFPESKLYWISSGNH	1764
QY	98	-----A-Y-R-----D-K-----	102	QY	246	-----D--GF--A-R-----I-----W-----T-K-DG-----	256
Db	687	SHRDIPTVKTVLWPNGLSLDIPAGRLYWDYFDRIETILLNGTDKIVYEGPELNHAF	746	Db	1765	TINRCLDGGLEVIDAMRSOLGKATALAIMGDKLWADQVSEKMGTCADGSGSVLVR	1824
QY	103	-----L-----AQQAA-----	109	QY	257	N--LA-----S--T--L--GO--	264
Db	747	GLCHGNLFWTEYRSGSVYRLRGVGGAPPTVLLRSRPPPIFEIRMYDAQQQVGTNK	806	Db	1825	NSTTLVHMVKYDESIOLDHKGTPNCPVNGDCSOLCLPTSTTRSCMCTAGYSRSGQ	1884
QY	110	AA-----A-A-----A-----A-RASO--	120	QY	265	-----HKG-----P-----I-F-A-----L--	272
Db	807	CRVNGGSSLCIATPSRQCAEQVLDADGVTCLANPSYVPPQCPQGFACANSRC	866	Db	1885	ACEGVGSFLLYSVHEGIRGIPLDPNKSDALVPVSGTSLAVGIDFHAENDTIYVDMGLS	1944
QY	121	Q-----G-----S-A-----K-----N-G-----E--	128	QY	273	-----K--W-----N-----K-----KGN-----P--I--L-----S--	283
Db	867	IQRWKCDGNDCLNDSDEAPALCHOHTCPSDRFKCENNRCPINRWLDCDNDGNSDE	926	Db	1945	TISRAKRDQWREDVVVTNGIGRVEGIAVDWIAAGNIYWTQDQPDVIEVARLNGSPRYVVIS	2004
QY	129	N-T-A-----N-----G-----E-----E--N--	136	QY	284	AGVDK-----T-----T-----I--I--I-----	292
Db	927	SNATCSARTCPNPQFSCASGRCPISWTCDLDDCGDRSDESASCAYPTCFLPTQTCNN	986	Db	2005	QGLDPRAITVHPEKGYLFWTEWGYPRIERSRLDGTERRVLNVVSI SWPNGISVDYQDG	2064
QY	137	G-----A--H-----T-----I-----	141	QY	293	--W-DA-----H-----	296
Db	987	GRCININWRCDNDGCDNSDEAGCSHSCSTQFKCNSGRCPPEHWTCTDGDNDGYSDE	1046	Db	2065	KLYWCARDTKLERIDLETGENREVLLSSNNMDFSVSVFEDFIYMSDRTHANGSIKRG	2124
QY	142	AN--N-----HTD-----M-----M--E-----V-D--	152	QY	297	-----TG-----B-----AK-----OQ-----	303
Db	1047	THANCTNQATRPGGCHTDFQCRLDGLCIPLRWRCGDGTDNDSSDEKSCGVTHVCDP	1106	Db	2125	KDNATDSVPLRTGIGVOLKDIKVFNRDRQKGTNVCANGCOQLCLYRGRQORACAH	2184
QY	153	G--D-----V-----E-I--PP-----NKAV-----V--	164	QY	304	-----F-----	304
Db	1107	SVRFCKDSARCISKAVCDGNDGDCDNDSDENSCSLACRPPSHPCANNTSVCLPPDKLC	1166	Db	2185	GMLAEDGASCREYAGVLLYSERTILKSHLSDERNLNAPVQFPEDPEHMKNVIALAFDYR	2244
QY	165	-----L--R-----G-H-----E--S-----E-----V--	172	QY	305	-----P-----F--H-----S-----A-----	309
Db	1167	DGNDGCGSDGEBELCDQCCLNNGGCSHNCVAPGEGIVCPCPLGMLGPDNHTCQISY	1226	Db	2245	AGTSPGTPNRIFFSDIHFGNIQIINDGSRRTIVENVGSGVEGLAYHRGWDLYWTSYTT	2304
QY	173	-----F-I-CA-----W-----N-----	178	QY	310	-----P-AL-----D-----V-D-----W-Q-----S--	318
Db	1227	CAXHLKCSQCKDNKFSVKSCYEGWVLEPDGHSRSLDPFKPFIIFSNNRHEIRRIDLHK	1286	Db	2305	STITRHTVDQTRPGAFERETVITMSGDDHPRAFVLDECONLMFWTNNEQHPSIMRAALS	2364
QY	179	-----P-----V-----S-----D-----LL--A--S-----	186	QY	319	--N-----N-----T--P--AS-----C-----S-----	326
Db	1287	GDYSVLVGLRNTIALDFHLSQALYWTDVVEDKIVRGKLLDNGALTSFEVWYQGLATP	1346	Db	2365	GANVLTLEBKDRTPNGLAIDHRAEKLYFSDATLKIERYDGHSHRYVLKSEPVHPFG	2424
QY	187	-----G-----S-----GD-----S-----T--	192	QY	327	-----TD-----M-----C-I--	331
Db	1347	EGLAVDWIAGNIYWNESNLQIEVAKLDGTLRTLLAGDIEHPRAIALDPRDGLFWTDW	1406	Db	2425	LAVYGEHFWTDWVRRAVQRANKHVGSNMKLLRVDIPQOPMGLIAVANDTNSCELSFCRI	2484
QY	193	A--RI-----W-N-----	197	QY	332	-----HV--CKLG--OD-----R-----	340
Db	1407	DASIPIREANMSGAGRTVHRETSGGWPNGLTVDYLEKRLWIDARSDAIYSARYDGS	1466	Db	2485	NNGGQDCLCLTHQGHVNCSC--GGRILQDDLTTCRAVNSSCRAQDEFCANGECINPSLT	2543
QY	198	-----LS-----E-----NST-----SG-----S--TQ--L--	209	QY	341	-----P-----P-----I--KT--	344
Db	1467	GHMEVLARGHFLSHPPFAVTLYGGEVYWDRTN--TLAKANKWTGHNVTVVQRTNPFDL	1525	Db	2544	CDGVPHCKDKSDEKPSYCNRRCKTFRQCSNGRCVSNMLWNCGADDCGDSDEI PCNKT	2603
QY	210	V-----L--R-----H-----C-----	214	QY	345	-----FQ-----	346
Db	1526	QVTHPSQPMAPNCEBANGGPGCSHLCLINYNRTVSCACPHLMKHLKONTTCYEPKFL	1585	Db	2604	ACGVGEFRCDGTICIGNSSRCNQFVDCEDASDEMNCSATDCSSYFRLGVKGLFQPCERT	2663
QY	215	-----RE-----	217	QY	347	-----G-----H--	348
				Db	2664	SLCYAPSWCDGANDCGYSDERDCPGRVPRCPLNYFACPSRCIPMSWTCDEKDDCEH	2723

```
QY 349 ---T-N---EV---N-A-I-K-W---D---PT---G 361
Db 2724 GEDETHCNKFCSEAQTECQNHRCISKQWLCGSDGDDCAHCEGKTCGSPSCFG 2783
QY 362 N---L---LAS-C---S--DD---M---371
Db 2784 THVCVPERWLCDDKDCADGADESIAAGLYNSTCDDREBFMCNQRCIPKHFVCDHDDC 2843
QY 372 ---T---LK---I---W---S--NK 379
Db 2844 AGSDSEPEPTCGPSEFRGANGCLSSRQWECGENDCHDQSDAPKNPHCTSPHK 2903
QY 380 ---QD---381
Db 2904 CNASSQFLCSSGRCAELLNCGDDGSDSBERGCHINECLSRKLSGCSQDCEDLKIGF 2963
QY 382 ---N---CV---H---385
Db 2964 KCRCPGFLKDDGRTCADVDECSITTFPCSQRCINTHGSYKCLCVGYPAPRGDPHSCKA 3023
QY 386 ---D---L-Q---Q---389
Db 3024 VTDEBFLIPANRYLRKLNLSNTLLKQGLNNAVALDFVREQMIYWTVDVTTQGSMI 3083
QY 390 ---H---N---K---EI---Y-TI---397
Db 3084 RMHLGNSNVQVLRHTGLSPDGLAVDWVGGNLYWCDKGRDTIEVSKLNGARTVLVSSG 3143
QY 398 --K-----W-----S-----P-----T-----402
Db 3144 LREPRALVVDVQNGYLTWDGSHSLIGRMGSSRSVVDTKITWPNGLTLVDYVTERI 3203
QY 403 -----G-----P-----CT 406
Db 3204 YWADAREVIEFASLDGSRHVVLSODIPIFALTIFEDYVYWTWETKSIINRAHKTCT 3263
QY 407 N-----N-P---N---ANL-MLAS-----A-S-P- 420
Db 3264 NKTLLISTLHRPMDLHVHALRQDPVPHPCNVKNGGCSNLCLL-SPGGHKACAPTNY 3322
QY 421 ---D---S-T---V---R---LW---DV-----D-----R-----431
Db 3323 LGSGRTCVSNCTASQFVCKNDKCIPIFWKDKTDDCGDSBDEPPDCPEFKCRPGQFQCS 3382
QY 432 -CIC-----IHT-L-----T-K-H-----Q-----442
Db 3383 TGICTNPATFICDNDNCDQNSDEANCDIHVCLPSPQFKCTNTRCIPGIFRCNGQDNCGDG 3442
QY 443 -----E-----P-VY-----S-----V-AF---450
Db 3443 EDERDCPEVTCAPNQFQCSITKRCIPRVWVCDRDNDVDSDEPANCTQMTGCVDEFCK 3502
QY 451 -S-----P-----DG-----R---Y-----LAS--G 460
Db 3503 DSGRCIPARWKCDGEDDGDGSDPEKPECDERTCEPYQFRCKNNRCPGRWCVDYDNDG 3562
QY 461 -----S-F-----D-K-CV-----466
Db 3563 DNSDEESCTPRPCSESEFSCANGRCIAGRWKCDGDHDCADGSDKCDKTPRCMDQFQCKS 3622
QY 467 -H-I---W-----NT-----471
Db 3623 GHCIPLRWKCDADCMGSDSEACGTGVRTCPLEDFQCNTLCKPLAWKCDGEDDGCN 3682
QY 472 -----QVCL-----H---476
Db 3683 SDENPECARFVCPNRPFRCKNDRVCLWIGRQCDGTDCGDTDEDCPEPTAHTTHCK 3742
QY 477 ---YL-----N-----479
Db 3743 DKKEFLCRNQRLSSLSRCLNMFDDCGDGSDEEDCSIDPKLTSCATNASTICGDEARCVTE 3802
```

```
QY 480 -----GQ-----V---L---L---NL--G-----R---S---IC-- 491
Db 3803 KAAACACRSGFHTVPQPGCQDINECLRFGTCSQLCNTKGGHLCSCARNFMKTHNTCKA 3862
QY 492 -----LY-----493
Db 3863 EGSEYQVLYIADNDNEIRSLFPGPHSAYBQAFQGDSEVRIDAMDVHVKAGRVVYTNMHTG 3922
QY 494 T---LP-----H-----HL-----V-----VI 502
Db 3923 TTSYRSLPPAAPPTTSNRHRRQIDRCVTHLNISGLKMPRGIAIDWAGNVYWTDSGRDVI 3982
QY 503 -----P-LVA--L-----IE-----L---LV-- 512
Db 3983 EVAQMKGENRKTLSGMDIDEPHALVVDPLRGTMYSWGDGNHPKPIETAAMDGTRETLVQD 4042
QY 513 -----L-----K 514
Db 4043 NIQWPTGLAVDYHNERLYWADAK 4065

RESULT 13
A53102
Alpha-2-macroglobulin receptor precursor - chicken
N:Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor;
C:Species: Gallus gallus (Chicken)
C:Date: 04-Sep-1998 #sequence_revision 04-Sep-1998 #text_change 09-Jul-2004
C:Accession: A53102
R:Nimphi, J.; Stifani, S.; Bilous, P.T.; Schneider, W.J.
J. Biol. Chem. 269, 212-219, 1994
A:Title: The somatic cell-specific low density lipoprotein receptor-related protein of t
A:Reference number: A53102; MUID:94103212; PMID:7506255
A:Accession: A53102
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-4543 <NIN>
A:Cross-references: UNIPROT:P98157; GB:X74904; NID:G438006; PIDN:CAAS2870.1; PID:G438007
C:Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated
d protein.
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding
C:Keywords: beta-hydroxyaspartic acid; beta-hydroxyaspartic acid; calcium binding; glycoprotein
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-3942/Domain: alpha-2-macroglobulin receptor 515K chain #status predicted <515K>
F:18-3942,3943-4543/Product: alpha-2-macroglobulin receptor #status predicted <MAT>
F:29-66/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:74-110/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:117-150/Domain: EGF homology <EG1>
F:156-190/Domain: EGF homology <EG2>
F:200-241/Domain: LDL receptor YWTD-containing repeat homology <YW01>
F:242-283/Domain: LDL receptor YWTD-containing repeat homology <YW02>
F:294-336/Domain: LDL receptor YWTD-containing repeat homology <YW03>
F:337-380/Domain: LDL receptor YWTD-containing repeat homology <YW04>
F:381-422/Domain: LDL receptor YWTD-containing repeat homology <YW05>
F:423-470/Domain: LDL receptor YWTD-containing repeat homology <YW06>
F:480-521/Domain: EGF homology <EG3>
F:573-615/Domain: LDL receptor YWTD-containing repeat homology <YW07>
F:616-661/Domain: LDL receptor YWTD-containing repeat homology <YW08>
F:662-712/Domain: LDL receptor YWTD-containing repeat homology <YW09>
F:713-754/Domain: LDL receptor YWTD-containing repeat homology <YW10>
F:755-797/Domain: LDL receptor YWTD-containing repeat homology <YW11>
F:805-840/Domain: EGF homology <EG4>
F:852-888/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:893-929/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:934-963/Domain: LDL receptor ligand-binding repeat homology <LDL5>
F:974-1009/Domain: LDL receptor ligand-binding repeat homology <LDL6>
F:1013-1049/Domain: LDL receptor ligand-binding repeat homology <LDL7>
F:1060-1095/Domain: LDL receptor ligand-binding repeat homology <LDL8>
F:1102-1138/Domain: LDL receptor ligand-binding repeat homology <LDL9>
F:1143-1180/Domain: LDL receptor ligand-binding repeat homology <LDL10>
F:1183-1219/Domain: EGF homology <EG5>
F:1225-1259/Domain: EGF homology <EG6>
F:1267-1306/Domain: LDL receptor YWTD-containing repeat homology <YW12>
F:1307-1353/Domain: LDL receptor YWTD-containing repeat homology <YW13>
```

F;1354-1396/Domain: LDL receptor WYTD-containing repeat homology <YW14>  
F;1397-1443/Domain: LDL receptor WYTD-containing repeat homology <YW15>  
F;1444-1486/Domain: LDL receptor WYTD-containing repeat homology <YW16>  
F;1487-1529/Domain: LDL receptor WYTD-containing repeat homology <YW17>  
F;1538-1576/Domain: EGF homology <EG7>  
F;1581-1624/Domain: LDL receptor WYTD-containing repeat homology <YW18>  
F;1625-1667/Domain: LDL receptor WYTD-containing repeat homology <YW19>  
F;1668-1711/Domain: LDL receptor WYTD-containing repeat homology <YW20>  
F;1712-1751/Domain: LDL receptor WYTD-containing repeat homology <YW21>  
F;1752-1794/Domain: LDL receptor WYTD-containing repeat homology <YW22>  
F;1795-1842/Domain: LDL receptor WYTD-containing repeat homology <YW23>  
F;1846-1882/Domain: EGF homology <EG8>  
F;1930-1972/Domain: LDL receptor WYTD-containing repeat homology <YW24>  
F;1973-2015/Domain: LDL receptor WYTD-containing repeat homology <YW25>  
F;2016-2059/Domain: LDL receptor WYTD-containing repeat homology <YW26>  
F;2060-2101/Domain: LDL receptor WYTD-containing repeat homology <YW27>  
F;2102-2147/Domain: LDL receptor WYTD-containing repeat homology <YW28>  
F;2155-2190/Domain: EGF homology <EG9>  
F;2195-2237/Domain: LDL receptor WYTD-containing repeat homology <YW29>  
F;2247-2288/Domain: LDL receptor WYTD-containing repeat homology <YW30>  
F;2338-2382/Domain: LDL receptor WYTD-containing repeat homology <YW31>  
F;2383-2423/Domain: LDL receptor WYTD-containing repeat homology <YW32>  
F;2424-2467/Domain: LDL receptor WYTD-containing repeat homology <YW33>  
F;2476-2511/Domain: EGF homology <EG10>  
F;2518-2555/Domain: LDL receptor ligand-binding repeat homology <LDLB>  
F;2560-2594/Domain: LDL receptor ligand-binding repeat homology <LDLC>  
F;2599-2633/Domain: LDL receptor ligand-binding repeat homology <LDLE>  
F;2646-2682/Domain: LDL receptor ligand-binding repeat homology <LDLE>  
F;2690-2724/Domain: LDL receptor ligand-binding repeat homology <LDLF>  
F;2732-2767/Domain: LDL receptor ligand-binding repeat homology <LDLF>  
F;2772-2810/Domain: LDL receptor ligand-binding repeat homology <LDLH>  
F;2816-2851/Domain: LDL receptor ligand-binding repeat homology <LDLI>  
F;2856-2895/Domain: LDL receptor ligand-binding repeat homology <LDLJ>  
F;2902-2936/Domain: LDL receptor ligand-binding repeat homology <LDLJ>  
F;2941-2977/Domain: EGF homology <EG11>  
F;2983-3018/Domain: EGF homology <EG12>  
F;3026-3065/Domain: LDL receptor WYTD-containing repeat homology <YW34>  
F;3066-3110/Domain: LDL receptor WYTD-containing repeat homology <YW35>  
F;3111-3153/Domain: LDL receptor WYTD-containing repeat homology <YW36>  
F;3154-3197/Domain: LDL receptor WYTD-containing repeat homology <YW37>  
F;3198-3238/Domain: LDL receptor WYTD-containing repeat homology <YW38>  
F;3239-3281/Domain: LDL receptor WYTD-containing repeat homology <YW39>  
F;3291-3327/Domain: EGF homology <EG13>  
F;3331-3366/Domain: LDL receptor ligand-binding repeat homology <LDLL>  
F;3371-3405/Domain: LDL receptor ligand-binding repeat homology <LDLM>  
F;3410-3445/Domain: LDL receptor ligand-binding repeat homology <LDLN>  
F;3450-3486/Domain: LDL receptor ligand-binding repeat homology <LDLO>  
F;3491-3528/Domain: LDL receptor ligand-binding repeat homology <LDLP>  
F;3533-3567/Domain: LDL receptor ligand-binding repeat homology <LDLP>  
F;3572-3606/Domain: LDL receptor ligand-binding repeat homology <LDLQ>  
F;3610-3644/Domain: LDL receptor ligand-binding repeat homology <LDLR>  
F;3651-3687/Domain: LDL receptor ligand-binding repeat homology <LDLS>  
F;3692-3728/Domain: LDL receptor ligand-binding repeat homology <LDLT>  
F;3738-3774/Domain: LDL receptor ligand-binding repeat homology <LDLU>  
F;3783-3820/Domain: EGF homology <EG14>  
F;3826-3858/Domain: EGF homology <EG15>  
F;3866-3909/Domain: LDL receptor WYTD-containing repeat homology <YW40>  
F;3910-3948/Domain: LDL receptor WYTD-containing repeat homology <YW41>  
F;3943-3983/Domain: alpha-2-macroglobulin receptor 85K chain #status predicted <EXT>  
F;3993-4011/Domain: 85K chain extracellular #status predicted <EXT>  
F;3996-4011/Domain: LDL receptor WYTD-containing repeat homology <YW42>  
F;4012-4055/Domain: LDL receptor WYTD-containing repeat homology <YW43>  
F;4056-4098/Domain: LDL receptor WYTD-containing repeat homology <YW44>  
F;4099-4141/Domain: LDL receptor WYTD-containing repeat homology <YW45>  
F;4150-4181/Domain: EGF homology <EG16>  
F;4199-4230/Domain: EGF homology <EG17>  
F;4235-4266/Domain: EGF homology <EG18>  
F;4271-4302/Domain: EGF homology <EG19>  
F;4307-4338/Domain: EGF homology <EG20>  
F;4343-4373/Domain: EGF homology <EG21>  
F;4376-4408/Domain: EGF homology <EG22>  
F;4421-4443/Domain: transmembrane #status predicted <TMM>  
F;4444-4543/Domain: intracellular #status predicted <INT>

F;116,138,187,276,359,448,731,926,1048,1152,1153,1193,1216,1305,1509,1556,1573,1614,1643  
3485,3559,3786,3837,3952,4074,4124,4178,4278/Binding site: carbohydrate (Aen) (covalent)  
F;188,2995/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F;2955/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted

Query Match 70.0%; Score 2556; DB 1; Length 4543;  
Best Local Similarity 10.1%; Pred. No. 2.9e-52;  
Matches 418; Conservative 73; Mismatches 17; Indels 3640; Gaps 366;

QY 1 M-----S-----IS-----SD 6  
DB 1 MGPELLAGCLLALLAAPAAALEAPKCTSPKQFACKDQITCISKGRWCDGKDCPDGSD 60  
QY 7 E-----V-----N-----F-----L 11  
DB 61 ESPDIPCQSKVSRQCNEHNCGLTGLCIHMSKLCNGLHDCFDGSDGPHCREQLANTAL 120  
QY 12 -----V-----Y-----R----- 14  
DB 121 GCQHCHCVPTLSGPACYCNNSFQLAEDRSCKDFDECTVYVTCSTCTNTEGSYTCSCVEG 180  
QY 15 Y-LQ-----E-----SG-----F--- 21  
DB 181 YLQDPNRSCKAKNEPVDPPVLLIANSONILATYLSGAPVPNITPTSAKQTTAMDENYI 240  
QY 22 --S---H---SAF-T-----F---GK-S-H-ISO-----S-N--- 38  
DB 241 EDTVCVWVGDSASQTLKCAKIPNLKGFVEERSINISLSHQVEMAWDMLTGNFYFVD 300  
QY 39 ---I---NG-----AL-----VP----- 45  
DB 301 DDDRIFVCNKGKGLTCVTLLELYNPKGIALDPAMKVFFTDYGOIPKVERCMDQNR 360  
QY 46 -----P-----A--LI-----SIQK-G-L-Q-Y--- 59  
DB 361 TKLVDKSVFPHGKITLDLSRLVYWDAYLDYEVVDYEGKRNHTIIQ-GILIEHLYGLT 419  
QY 60 V-E-----A--EVS-I---N--E-----D-G-----T--- 71  
DB 420 VFENLYATNSDNANAQKTSIVRNRFNSTEVQVTVRVKDGALHYHORRQPTVRSHA 479  
QY 72 -----L-----F---DG----- 75  
DB 480 CEPDQFGKGGCDICLLGNHSHKSRCTCRSGFSLGSDGSKCKKPELFLVYKGRPGI 539  
QY 76 -R-----PIE-----SL-----S-LI----- 84  
DB 540 IRGMDMGAKVPDEHMTPIENLMPALDFAETGFIYFADTTSYLGKIDGTERETIL 599  
QY 85 -D-----AV--M-----PD--V-V-----QTR----- 95  
DB 600 KGIHNVGIIADVDMGNLWYTDGPKKTSIVARLEKAAQTRKTLIEGKWTHTPRAIVVDP 659  
QY 96 -----QQA----- 98  
DB 660 LNWGMYWTDWEEDPKDKRGKITERAWMGDSNRNVFTTSKTLVLPNGLSLDIPAKILYWVD 719  
QY 99 -----YR-D---K--- 102  
DB 720 AFYDRIEMVYLNGTERKIVYEGPELHAFGLCHYSSFLFWTEYRSGSIYRLDQSSKAVSL 779  
QY 103 L-----AQQA-----AA-----A-A-----AAAA 117  
DB 780 LNERPPIFEIRMYDAQOQVGSNCRVNVNGGSSLCLATPRGRQCAQADQLIGADSVT 839  
QY 118 --A--S---Q-Q-G-----S--A----- 124  
DB 840 CEANPSYIPPPQCPQGEFAKKNRCKIQRKWKCDGNDCLDNDSEAPELCHQHTCPSPDRFK 899  
QY 125 -KN-----G---E-N-T--A-----N----- 132  
DB 900 CKNRRCIPNRWLCDGNDGDNCDNNEDESNSTCSARTCSPNFSCASGRCPISWTCDLDDDC 959





Db 3116 CDKGRDITVSKLNGAIVRTVLVNSGLRPREALVVDVQVNGVLYWTDWGDHSLGKIGMDGT 3175  
QY 400 --S--P--T--G--P-----404  
Db 3176 NRSVIVDTKITWPNGLTLOYINSRIYWADAREDIYEFASLDGSGNRHTVLSQDIPHFALT 3235  
QY 405 -----GTN-----N-P---N 410  
Db 3236 LFEDYIYTDWETKSIINRAHKTGTANKLLISTLHRPMDIHIVHPYQDPVPHNPKCTNN 3295  
QY 411 A---NL-MLAS-----A-S-F--D---S-T---V---R---LW---DV---429  
Db 3296 AGCSNLCLL-SPGGGHKACPTNLYLSDGKTCVSNCTASQFVCKNDKICIPFWKCKDTE 3354  
QY 430 ---DR---GIC-----IHT-L---438  
Db 3355 DCDGRDEPDCPEFKRPFQFQFCSTGICTNPAFICDGDNDQDNDSEANCDIHVCLPSQ 3414  
QY 439 ---T-K-H---Q-----E-----P-VY-----446  
Db 3415 FKCTNTRCIPGIFRCNGQDNCGDGDEKDCPEVTCAPNQFQCAITKRCIPRVWVCDRON 3474  
QY 447 ---S-----V-AF--S---P-----DG-----R---455  
Db 3475 DCDVGSDEPANCQTMTGVDPEFRCKDSGRICIPARWKCDGDDCGDSEPKEDBERTCE 3534  
QY 456 -Y-----LAS-G-----S-F-----462  
Db 3535 PYQFRCKNNRCVPRWQDYDNDGNSDBESCTPRPCESEFSCANGRCIAGRWKCDGD 3594  
QY 463 ---D-K-CV-----H-I--W-----Q-----469  
Db 3595 HDCADGSDKDCIPRCEFDQYQCKNGHCIPMRWRCDADCDMDGTDEDCGTGVRTCPLD 3654  
QY 470 ---NT-----Q-----VCL-----475  
Db 3655 EFCQNTLRKPLAWKCDGEDDGDNDSDENPECLKFCQPNRFRCKNDRVCLWIGRQCD 3714  
QY 476 -----N-----G-QVLL-----N-L-----G 487  
Db 3835 LCNNTKGSHVCSAQNFMKTDNMCKAEGSHQILYIADDNKIRSMYFPNPSAYEPAFQ 3894  
QY 488 ---R-----S-I-C-----491  
Db 3895 DENVRIDAMDIYVKGNIYWNHTGRISYCELPASSAASTANRRNRQIDGGVTHLNTS 3954  
QY 492 -L-----Y-----TL-----PH-----497  
Db 3955 GLKMPRGIAVDWAGNIYWTDSGRDVEIVQAQMGKGNKTLISGMIDEPHAIIVDPLRGTM 4014  
QY 498 ---H-----LV-----VI-----502  
Db 4015 YMSDGNHPKIETAAMDGTILRETLVQDNIQWPTGLAVDYHNERLYWADAKLSVIGSIRLN 4074  
QY 503 ---PLVA---L-----I-E-----L---L-----511  
Db 4075 GTPVPAIDNKKGLSHPPFSIDIFEDYIYGVTYNNRIFKHKFGHKSVTNLTSGLNHATD 4134  
QY 512 -VL-----K 514  
Db 4135 VLYHQYK 4142

RESULT 14  
A47437

LDL-receptor-related protein - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: A47437; S27801; T21547  
R:Yochem, J.; Greenwald, I.  
Proc. Natl. Acad. Sci. U.S.A. 90, 4572-4576, 1993  
A:Title: A gene for a low density lipoprotein receptor-related protein in the nematode C.elegans  
A:Reference number: A47437; MUID:93281621; PMID:8506301  
A:Accession: A47437  
A:Molecule type: DNA  
A:Residues: 1-4753 <YOC>  
A:Cross-references: UNIPROT:Q04833; GB:M96150; NID:G156359; PIDN:AAA28105.1; PID:G156360  
A>Note: nucleotide sequence not given; translation not complete in this paper  
R:Yochem, J.; Greenwald, I.  
submitted to the EMBL Data Library, July 1992  
A:Description: A gene for an LDL receptor-related protein (LPR) in the nematode C.elegans  
A:Reference number: S27801  
A:Accession: S27801  
A:Molecule type: DNA  
A:Residues: 1-4753 <YOC>  
A:Cross-references: EMBL:M96150; NID:G156359; PIDN:AAA28105.1; PID:G156360  
R:Wilkinson, J.  
submitted to the EMBL Data Library, June 1996  
A:Reference number: Z19439  
A:Accession: T21547  
A:Status: preliminary; translated from GB/EMBL/DBDJ  
A:Molecule type: DNA  
A:Residues: 1-4753 <WIL>  
A:Cross-references: EMBL:273907; PIDN:CAA98124.1; GSPDB:GN00019; CESP:F29D11.1  
A:Experimental source: clone F29D11  
C:Genetics:  
A:Gene: LPR  
A:Map position: 1  
A:Introns: 3/1; 88/1; 132/1; 172/3; 219/1; 298/1; 463/2; 526/2; 585/3; 780/2; 874/2; 975/15/1  
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding  
C:Keywords: tandem repeat; transmembrane protein  
F:53-87/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F:92-131/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F:138-175/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
F:182-218/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
F:223-257/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
F:262-297/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
F:302-336/Domain: EGF homology <EGF1>  
F:1054-1095/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
F:1101-1138/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F:1146-1182/Domain: LDL receptor ligand-binding repeat homology <LDL9>  
F:1187-1223/Domain: LDL receptor ligand-binding repeat homology <LDL10>  
F:1228-1263/Domain: LDL receptor ligand-binding repeat homology <LDL11>  
F:1270-1307/Domain: LDL receptor ligand-binding repeat homology <LDL12>  
F:1313-1350/Domain: LDL receptor ligand-binding repeat homology <LDL13>  
F:1359-1396/Domain: LDL receptor ligand-binding repeat homology <LDL14>  
F:1441-1475/Domain: EGF homology <EGF>  
F:1611-1654/Domain: LDL receptor YWTD-containing repeat homology <YW33>  
F:2792-2829/Domain: LDL receptor ligand-binding repeat homology <LDL15>  
F:2834-2868/Domain: LDL receptor ligand-binding repeat homology <LDL16>  
F:2874-2912/Domain: LDL receptor ligand-binding repeat homology <LDL17>  
F:2919-2956/Domain: LDL receptor ligand-binding repeat homology <LDL18>  
F:2961-2997/Domain: LDL receptor ligand-binding repeat homology <LDL19>  
F:3006-3044/Domain: LDL receptor ligand-binding repeat homology <LDL20>  
F:3049-3093/Domain: LDL receptor ligand-binding repeat homology <LDL21>  
F:3100-3135/Domain: LDL receptor ligand-binding repeat homology <LDL22>  
F:3140-3174/Domain: LDL receptor ligand-binding repeat homology <LDL23>  
F:3187-3222/Domain: LDL receptor ligand-binding repeat homology <LDL24>  
F:3586-3623/Domain: EGF homology <EGX1>  
F:3627-3666/Domain: LDL receptor ligand-binding repeat homology <LDL25>  
F:3671-3705/Domain: LDL receptor ligand-binding repeat homology <LDL26>  
F:3709-3746/Domain: LDL receptor ligand-binding repeat homology <LDL27>  
F:3753-3788/Domain: LDL receptor ligand-binding repeat homology <LDL28>  
F:3793-3830/Domain: LDL receptor ligand-binding repeat homology <LDL29>  
F:3833-3871/Domain: LDL receptor ligand-binding repeat homology <LDL30>  
F:3878-3912/Domain: LDL receptor ligand-binding repeat homology <LDL31>  
F:3917-3951/Domain: LDL receptor ligand-binding repeat homology <LDL32>

F:3959-3995/Domain: LDL receptor ligand-binding repeat homology <LD33>  
F:4000-4040/Domain: LDL receptor ligand-binding repeat homology <LD34>  
F:4049-4083/Domain: LDL receptor ligand-binding repeat homology <LD35>  
F:4092-4130/Domain: EGF homology <EGF2>  
F:4343-4386/Domain: LDL receptor WYTD-containing repeat homology <YW38>

Query Match 70.0%; Score 2553.7; DB 1; Length 4753;  
Best Local Similarity 9.8%; Pred. No. 3.7e-52;  
Matches 428; Conservative 70; Mismatches 10; Indels 3853; Gaps 375;

```
QY 1 MSIS-----SDE-----V-----N-PL-----V-----12
Db 105 HVSCTAEYORCDNITDCADGSDKDCPASTVDCSSQNVFMCADGQCQFDVSKKCDGKYD 164
QY 13-----Y-----Y-----13
Db 165 CRDLSDEKDCSRNHTACQYQPRCADTKQCIQKSWVCDGSKDCADGSDPEPTCEPKCT 224
QY 14-----R-----Y-----15
Db 225 ANEFQCKNRCQPRKRCRDYDDCGSDNDECEGYRCPPGKWCNCPGTGHCIDQLKCDG 284
QY 16-----L-Q-----E-----SG-----FSH-----S-----24
Db 285 SKDCADGADQCCSONLCPSLGCQAGCPSPHGECTCFSGYKLDLDRF-HRTCSDINECA 343
QY 25-----A-----FT-----K-----SH-----I-----27
Db 344 EFGYCDQLCANHRPGFTCSCLGDCFTLOMEHGGKDNLTWRGVCVSNADKMLFVARRE 403
QY 28-----P-GI-----K-----L-----SH-----I-----36
Db 404 GLYRLNPKNPDEEVKGLASGEFTYGDYDGRKIEFTWDLRAHSAFSDVDEGEISQIK 463
QY 37-----S-----N-----I-----39
Db 464 KLSKSLVPRCLAVDWNITNTLYIESGRRIDVSSYDGERRTVLLADGLTLPDLALDP 523
QY 40-----N-----G-----A-----L-----43
Db 524 LRGEFTNQLKLEAAMDGTNRRLTVNTHQVSGIVVDITAKRVYVVDPKVDRLESID 583
QY 44-----VP-P-A-----A-L-I-----S-----I-52
Db 584 YQGNDRRIVAQGNVVPHPFGLAPQLYLWTDWTLGLVQVEKFGSDTKLLWSNTENN 643
QY 53-----I-----Q-----K-----GL-----Q-58
Db 644 FPMGISAYHPMAQPGQSECLAMKIENPCTNADCEGMCILSKDNGGFGVGYKCAPIGQ 703
QY 59-----Y-----V-----EA-----EV-----64
Db 704 KLVNGKRCIDSIDYLLFPSSNKIVRGIFPEINEKALAEAVLPISPSQRRIGMVEVECDV 763
QY 65-----S-I-N-----E-----DG-----70
Db 764 HGNSFFYADIMNTIYIRPDGEGAAPLVTHNDGLFSMSFOWISKQLYYVDNIRNSLEV 823
QY 71-----T-----LF-----D-----74
Db 824 VKIGETGLVHPDELVRRLITELRDPVSVVHPWKGLLFYAEAMRPAAYRCHIDQNCQ 883
QY 75-----GRP-I-E-----SL-S-----L-ID-A-VM--88
Db 884 VIRNTTLGRPSEMAIDFAENRLCWGDTLLKTIKSCMPDGNVVKLDINPIPAITIME 943
QY 89-----P-----D-V-----V-----92
Db 944 YIYVYHQPYSIRVHKNGGSGKIYREFCADERSIFSLKACSHQNPQIPDDSRHPCPRA 1003
QY 93-----Q-T-----RQ-----Q-A-----Y--99
Db 1004 SOCTQLCFATPSHSHNELEAKCACRQGFWMINKENNHSQKQPAEKIEQLCSNSNQFQC 1063
```

```
QY 100---R-----DK-----L-AQ-----105
Db 1064 KNGRCIPKWKCDGENDCLDESDEIDKDGKCFHETECAENTIKCRNTKKCIPAQYCGDG 1123
QY 106-----Q-----Q-----107
Db 1124 DNDGSDYDESDYKCKGQKPVCAAKKFOCDNHRCIPEQWKCDSDNDGSDGSEKLEMG 1183
QY 108-A-AA-----A-----A-----A-----113
Db 1184 NATCAHQSCANGRCIPIYWLCDGNDYDGTDEDKERCPPVQCSALQFCANGRCQVCP 1243
QY 114-----AAAA-----AS-----QO-----GS--123
Db 1244 LRNHCDGSDGSDGSDSCAVTAESCTPDQFKCVSSGLCIPASWKCDGQDQDCDSDG 1303
QY 124-----A-K-----N-----GEN-----T-----130
Db 1304 KFGCTSGRQSSDPQFKCGNGRCILANNWLCGSDGSDGSDGSESGCKTSMNARKCPFEH 1363
QY 131-A-N-----GE-----E-----NG-----137
Db 1364 VACENDQETCIPHLQLCDGKTHCPGTDGGRCARDLCSADRAGCSFKCHNSPENGICSC 1423
QY 138-----AH-----T-IA-----142
Db 1424 PFGEQLVNTKCEPENECILSSSCQRCDEKHGFTCSDEGYELVDVKRTCKVADNVKD 1483
QY 143-----N-----N-HT-----DMWE-----150
Db 1484 TRIYVSNRRIYSDHKLNDWHTFGAIVENAJALAWDSLTDRIYSDIREKKILSANRNG 1543
QY 151---V---DG-DV-E-----I-----P--158
Db 1544 TNATFIADGLDITGIALDWGRNLYWVDSLNTIEVANLEDPKORTLLVHQNVSQPRG 1603
QY 159---P---N---KA-----VV-----L---R-----166
Db 1604 IAVDRKGMVFWTDGQNCPIERASMDGTDRQIIIVTKIYWPNTIALDVTTRVYFADSK 1663
QY 167-----G-----H-----E-----SE 171
Db 1664 LDFIDFVNYDGSRTQVLASSKFVQHPHALAIFEDMMYISDRRLQKLQVYPKYPNGTTSE 1723
QY 172---VF-----I-----174
Db 1724 YPSHTFSKALGVVAVHPVLQPVIKNNPCSTNPCSHLCLLNKNNTFTCKCPMGEKLDASGK 1783
QY 175-C---A-----W-N-----PV-----S-----D-----182
Db 1784 KCIDDAKPLVIIQKTNVFGIEMNSASEKETPVLAGMVPLSLGLGNADFADYDALSEEMFI 1843
QY 183 L-----LA-----S-----G-----G-D-----190
Db 1844 LEHTNHAKTLAGIQTDDSAIYRSTVNGNKTKMFSSAVPDADAYCLGFDWNGRNLVVGNKIT 1903
QY 191-S-----TA-----R-----I-----W-----196
Db 1904 QTIEIRTOGKQYRSVILNSDOSPTAVTVPVIAVDADKGYVFWLDRGGGAADAKVARAG 1963
QY 197---N-L-S-E-----N-----S-T-----S-----204
Db 1964 LDGSPNPLVIASNDLAEHLDAIDTNTVYVFSKAGRISSVYDQDRHYVLSDGGRQP 2023
QY 205-----G-----S-----T-----Q-----L-V-L-R-----212
Db 2024 NGLAFYDRLFYADSAFDSIEVATINGDSQPQWTHFKKDVENLANIKALQPRASSGHP 2083
QY 213---H-CI-----R-----E-G-----CO-----D--221
Db 2084 CHINNGNCNDHICILPMFAORTCTCANGYVKDGTCKLFDSEFVIVATKTKVIGYPIDET 2143
```

QY 222 ---V---P---S---NK-----D-V--- 228  
Db 2144 QSKGVAMEPIGGLSITGVYDYESKIYVAASGINKGITAYIGESSPRAVIRDSIGSL 2203  
QY 229 T--SL--DW-----N---S--E--GT---LL----- 240  
Db 2204 TIKSLAIDWINYMYFINHAERTNIEVSKLDGYRKILLTTKTETPSSIAVDVPSRYLY 2263  
QY 241 -A-----T-----GS-----Y-----DG----- 247  
Db 2264 WADQOQPTTQSRFLDGSREVIVSSGIAEPTDLVVDVASKMYWSDAKMDGIYRVRSTG 2323  
QY 248 -----FAR--I-----WT-----KD-G 256  
Db 2324 GTPELVRSDIASAAGVALHGQNYWTDNRLEKFRATSKPNQTSLLSPPTVAASLKDIG 2383  
QY 257 -----NL----- 258  
Db 2384 DVAVFSSNNQPRASSPCQITDNLKSPCTQLCFATPTCTCARGVLKGRTCBEPDT 2443  
QY 259 -----AS-----T----- 261  
Db 2444 YIMFSDGDKIIDVAIBPVKASRLKDPPEISNLQTFDVLNRRVYFVESPVGNIS 2503  
QY 262 -----L-G-----O-----H- 265  
Db 2504 WFSMNAENPRLVFGASKQPHAKEIRHISDMKLDWLTQKIYFTTGRGKVMADTAGHEHL 2563  
QY 266 -----F-----A---L-----K-----W----- 274  
Db 2564 STIASGDWTYALAI DPCSGLLFWSDGYKTSGLYEPRIERSNLAGSRKVI VSESISLP 2623  
QY 270 -----F-----A---L-----K-----W----- 274  
Db 2624 AAI AVDFRNQKIYADVNRLNIEVADYDGNRKRVIASGYRAKSLDIWRMLYMSDPLSNG 2683  
QY 275 -----N----- 275  
Db 2684 VFRIDKESGLENVSDRRIPGALRVFASESDVTRNQVNCNALTSQLCTDNGCQDL 2743  
QY 276 -----K-----K-----G-----NF----- 280  
Db 2744 TVVADDIGLAASKVQCSNDTYELVQEPGKYPTQCVLRGSNSEPAKECLPPYFQCGDG 2803  
QY 281 --IL----- 282  
Db 2804 SCILLGATCDKPCADASENPYCNTRSCPEDYNLCTNRRCIDSARKNHIDDCGDS 2863  
QY 283 -----SA-----G-----V-DK--T----- 289  
Db 2864 DELDCPSAVACABGTFFCSNGHCINQTKVCDGHNDCHDQVSDSLATCPGLPIDCRGVK 2923  
QY 290 -----T---I----- 292  
Db 2924 VRCPTNINICQPADLCGYDDCGKADENOLFMCNQCAQHYVRCPSGRICIPETWQCDGD 2983  
QY 293 -----WD-AHT-----GAK-----Q-F----- 304  
Db 2984 NDCSGDWDETHNTCTDTAG-KICVGDYLFQCDNLKCSIRAFICTDGEDDCGDSDEHSRH 3042  
QY 305 -----PPH--S--A-----P-A-L--DV-----D----- 315  
Db 3043 GCGNRTCTQEFHCTNAKLAQKYECIPRAWLDCDGDVTCAGGEDSBSTELCKTEKCKCNK 3102  
QY 316 -----W-----Q----- 317  
Db 3103 GEPRCSNQHCITWECDDGNDCLDGSDEHANTYSSCQDPDFQCANHKVPSNWKCDGN 3162  
QY 318 -----S-----N-----N-----T-----F-----AS-- 324  
Db 3163 DDCEDGSEDKCPKNSAQAASKCSNGQPQCTSGECIDDAKYCDRNFDCDRSDESLC 3222  
QY 325 ---CS---TDM-----C-----I-----H-V-----C- 334

Db 3223 FIDCSLAEPICEQCKMDMKIGYKCDCEFAIDISDQKSCHNVNCEYEGISGCSQKCD 3282  
QY 335 -KLG-----Q---D---R---P-----I-K-T-----F- 345  
Db 3283 DKIGSYKCGVDGYQLSSDDHSCRKTEPEPFLLANKHYIRKISIDGNKYELAAQGF 3342  
QY 346 -----OG-----HTN-----E-----V- 352  
Db 3343 NVVSLDIDLTEKAYLIDQGLRLRLRVLDLDEMDSPSSYETVLRH-NVYTEGIAVDWVG 3401  
QY 353 -----N-----AI-----K-----W----- 357  
Db 3402 RKLMLNROERSIRVCELDGRFCKTLIRDRIOQPAIVVHPGKGYLFFTEWSLOPVIGRI 3461  
QY 358 -----DP-----TG-N-----L-----A 365  
Db 3462 ALDGPSELQDPIFKLAHDLGWPNAIAIDYFSDRLFWGDAHLNEIGFMDFGNRRHIPA 3521  
QY 366 -----SCS---DD-M---TL-KI---WS-----MKQ- 380  
Db 3522 QRTSHVS-SMVVFDLYLWADNLRREVLCRWKGTGKNETILKTVLPNDLRIVHPMRQP 3580  
QY 381 -----DN---CVH-----DLQ-----Q-----HN-KE-IYT----- 396  
Db 3581 AYPNPGDNNGCSHCLIGAGNGYTCSPD-QPVLLSDQKTCENCTERQFACGGDDA 3639  
QY 397 --I-K-W-----S-P-T-G---P-G---TN-N---P-----N----- 410  
Db 3640 KCIPLMYCDGBCDRCDSDEPESICGQRI CPVGBFQCTNHNCTRPFOICDGNDDCGDS 3699  
QY 411 ---ANL-----ML---A-----S-A-----S-----F- 420  
Db 3700 SDEQNCDAKADPWFKAATGRICIPRRFTCDGDDGDRSDEADTLCMSAERNCTAEFR 3759  
QY 421 -----D-S--T---V-----R---L---W----- 427  
Db 3760 CANNKCIKAWCRDNDGCGSDETPCAQIECKGWTRCSCSSYRCIPNNAFCNGQDDC 3819  
QY 428 -----DV-----D----- 430  
Db 3820 RDNSEDKQRCPTCDDVGEFRCATSGKCI PRRWCMCTENDCGDNSDELASCGTTRPCS 3879  
QY 431 ---R-G-I---C---I-----H-----T-LYGH---QE 443  
Db 3880 ESEFRNDGKCI PGSKVCDGTIQCSDGLDESQCTLRRLCPGHRQCDGTGCI AEHKWCDRK 3939  
QY 444 -----P-----VVS---VAF---S-P---DG----- 454  
Db 3940 KDCPNADELHCEDEV-SRRTCSPPFEPCANSVICI PRKFMCDGNDGDNDSSTSECSA 3998  
QY 455 -----RY-LA-S-----G-----SF-----D----- 463  
Db 3999 QCDPPLRFRCAHSRLCLNLIQLCNGFNDCGPNDFSDHLSMCSSEFSEYGCSSDQFKCAN 4058  
QY 464 -KCV----- 469  
Db 4059 GKCVNGTVACDRKDCGDASDEIGCSKHGKTSCEAFNGNGGCKHICTDVRDGFYCHCRD 4118  
QY 470 -----N--TQ-----V-----C----- 474  
Db 4119 GFRPDQSPKECIDDECAENNTCTQLCLNTKGSYLRCHEYYNNVVGSMTGKDCRAK 4178  
QY 475 -----LH-----Y----- 477  
Db 4179 GDAANVMIGADDSLVQLSLHSGTNRHAAKANDNDIIGIAFDPKELMYWIDGSRT 4238  
QY 478 -L-----N---CO-----V-----LL-----NL-G----- 487  
Db 4239 IYRSIAIANGSHQEGKLDVDFAAMGVPTAIVDYTTGNLFLIAAVSENIENGLVTARKK 4298  
QY 488 R-S-----I-CL---Y-----T-L--P-----H----- 497



QY 197 -----NL--SE-----NS--TS----- 204  
Db 2775 INRGTSQSCMGGAIOCNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 2834  
QY 205 -----GS-----TQ--LV----- 210  
Db 2835 PSCSDPDGHCESSTKAPACKGVCCEPDYVYLNKNCVPRIECGCKDTQGVLIIPADKTM 2894  
QY 211 L-----R-----HC-----I----- 216  
Db 2895 INRGTSQSCMGGAIOCNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 2954  
QY 217 -----EG--G-----Q-----D-----VP-----SNKD--VTSL--D- 232  
Db 2955 PSCLDSEGHCEGTTKAPACQGGCVCEPDYVYLNKNCVPRIECGCKDAQGV--LIPADK 3012  
QY 233 -W-N-----SE-----G-----T-L-L--A----- 241  
Db 3013 TWINRGTSQSCMGGAIOCNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 3072  
QY 242 -----T-----GS-----Y-----D-----G-F-AR--I----- 251  
Db 3073 QPSCSNTDVHCEGSPNTLLSCREGCVQSGYVYLNKNCVPRIECGCKDAQGVLIIPADK 3132  
QY 252 -W-----T-----KDG-----N--L--A-S--T----- 261  
Db 3133 TWITSGTSQSCMGGAIOCNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 3192  
QY 262 -----L-----G-----Q-----HKGP-----IF-----AL-----K 273  
Db 3193 CPPLCLDPGLCBGTSKVPSTCREGICQPGYVYLNKNCVPRIECGCKDAQGVLIIPADK 3252  
QY 274 -W-----N-K-----K-----G-----N-FI--L-- 282  
Db 3253 TWISRGTSQSCMGGAIOCNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 3312  
QY 283 -----SA-----G-V--DK--T-----TII--W- 293  
Db 3313 CVFSCNRCEVTSPVSSCREGLCNHGFVFBEDKCVFRTQCGCKDAQGVLIIPADK 3372  
QY 294 -----D-----AHT----- 297  
Db 3373 SKGTSQSCMGGAIOCNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 3432  
QY 298 -----G-----E-----A-----K-Q----- 302  
Db 3433 SCLDPEGLCKDISPKVPSTCKGCVQSGYVYLNKNCVPRIECGCKDAQGVLIIPADK 3492  
QY 303 -----Q-----F-----P--F-----HS----- 308  
Db 3493 SPGTSQSCMGGAIOCNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 3552  
QY 309 -----A-----PA-----L-----D-----V--D--W- 316  
Db 3553 SCLDPEGLCKGASPKVPSTCKGICQSGYVYLNKNCVPRIECGCKDAQGVLIIPADK 3612  
QY 317 -----Q-----SN-----N-----T--F--A 323  
Db 3613 SRGTSQSCMGGAIOCNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 3672  
QY 324 SC-----ST-----D-----D----- 328  
Db 3673 SCLDPEGLCKGASPKVPSTCKGICQSGYVYLNKNCVPRIECGCKDAQGVLIIPADK 3732  
QY 329 -----M-----C--I-----H----- 332  
Db 3733 SPGTSQSCMGGAIOCNFKCPSEAYQODMEDGNSNCTSIPLQCPAHSHTNCLPTCQ 3792  
QY 333 -----VCKL-G--Q-----D-----R-----P----- 341  
Db 3793 SCLDPSALCGTSKVPSTCKGCVQSGYVYLNKNCVPRIECGCKDAQGVLIIPADK 3852

QY 342 -----I-----K-T-----FQ--G-----H-T-----N 350  
Db 3853 STGCIQSCACVEGTTIQCNFQCPPTGTYCNHNNCAKIPLOCPAHSHTNCLPTCQ 3912  
QY 351 -----E-----V-----N-----A-----I--K-W--D- 358  
Db 3913 LDGSCQTSKVPSTCKGICQPGYVYLNKNCVPRIECGCKDAQGVLIIPADK 3972  
QY 359 -----P--T-----G-N-----L-----L--A- 365  
Db 3973 TOSCACTGAGVQCNFQCPPLGTYCKDSGDSNCTKIHKGAMGDVLMAGGIRALQCPAH 4032  
QY 366 -----SCS--D-----SCS--D-----DM--TL-----K-- 374  
Db 4033 SHFTSCLPSCPSCSNLDGSCVESNFKAPSVCKGICQPGYVYLNKNCVPRIECGCKDT 4092  
QY 375 -----I-----W-----S-M-----K-----Q----- 380  
Db 4093 QGGLIPAGRTWISSDCTKSCMGIIQCRDQCPPTGYCKESNDSSRTCAKIPLOCPAH 4152  
QY 381 -----D-----D-----N--CV--H--D- 386  
Db 4153 SHYTNCPLPACSRCTDLGHCEGTSKVPSPCKEGCLCQPGYVYLNKNCVPRIECGCKDA 4212  
QY 387 -----L-----Q-----Q-HN-----K----- 392  
Db 4213 QGGFVPAGRTWISRGTSQSCACVGAOCHNFTCTGTQCNSSCKITVQCPAHSHTYTT 4272  
QY 393 -----E--I--Y-----T-----I 397  
Db 4273 CLPSCLPSCDFPEGLCGASPRAPSTCEGVCCEADYVLRDKCVLRTOCGCKDAQGVLI 4332  
QY 398 -----K-----W-----S-----P----- 401  
Db 4333 PANKTWLRGCAQKCTCKGNIHCNFKPLGTECKDSVDGNSNCTKIALQCPAHSHTY 4392  
QY 402 -----T-----G-----P-----GT-- 406  
Db 4393 CLPSCIPSCSNVDRCESTSLQRPSTCEGCLCHSGFVSKDKCVFRTQCGCKDAQGVLI 4452  
QY 407 -----N-----P-----NAN-----L-M 414  
Db 4453 PAKGNIWTTGSCORCTCTGGLVQCHDFQCPSEABQDIEDGNSNCTSIPLQCPAHSHTY 4512  
QY 415 -L-----A-----S-----A----- 418  
Db 4513 CLPQCPSCSDPDGHCESSTKAPACKGVCCEPDYVYLNKNCVPRIECGCKDAQGVLI 4572  
QY 419 -----S-----P----- 420  
Db 4573 PESKTVWSRGCTKNCCTCKGTVQCHDFQCPSEABQDIEDGNSNCTSIPLQCPAHSHTY 4632  
QY 421 -----D-----D-----ST-----V-----R-----L- 426  
Db 4633 CLPSCLPSCDFPEGLCGTSPEAPSTCEGVCCEPDYVYLNKNCVPRIECGCKDAQGVLI 4692  
QY 427 -----W----- 427  
Db 4693 KTGQTVISSGCKICTCKGFFQCPQSKYKCPSTQCESEDSGNSNCTSIPLQCPAHSHTY 4752  
QY 428 -----D-----D-----D--V-----D-RG-- 432  
Db 4753 CLPCLPSCSNPDGRCEGTSKAPSTCEGVCCEPDYVYLNKNCVPRIECGCKDAQGVLI 4812  
QY 433 -----I-----C--I-----H-----T----- 437  
Db 4813 PAGNTWISSDCTKSCMGIIQCRDQCPPTGYCKESNDSSRTCAKIPLOCPAH 4872  
QY 438 -LT-----K-----H----- 441  
Db 4873 YLTFDGTFTYFLGRMNYLKITVDKLPRIEPLIMEGRNKISPKGSSTLHEVTIVGYK 4932  
QY 442 -----QE-----PVYS-----V--A-----F-----S-P- 452

4933	IQOEELVVLVNDKVAVP-YNPNEHLRVMLRAQRLLLVTDFFEMVLDFDGKHSAVISLPT	4991
453	--DG--R--Y--LAS--G-S--P--	462
4992	TYRGLTRGLCGYDRDOSNELMLPSGVLTSNVHVFGNSWEVKAQHAFFRPFPALPEDEER	5051
463	DK- C-V--H--	467
5052	DEEPLLQSECQSOTALISSTQACRVLVDPPGFPAACHQIIAPEPEQRCMLDMCTGWK	5111
468	--I--W--NT-Q--	472
5112	TKEEELRCRVLSGYALICQEGANMTGWRDHTHCAMTCTPANTVYQRCMTPCPACAKFV	5171
473	--VC--L--H--Y-L--N--G	480
5172	TPKVCGEPCVEGCASLPGYIYSDTSQSLPVTHCGCTADGIYYKLGDSPVNTDCSQHCTCAS	5231
481	Q-VLLN--LG-R--SICL--YT--L--P-H--H--	498
5232	QGILLCEPYCRAGES-CWVANFTRGCFQDSPCLQNPCHNDGRCEBQATFICHDGPGYG	5290
499	-----LWVI-P-LV-ALI-	509
5291	GEFCTEPQDITTRKIEASSLVAILPGVLVWMLVPVLLPRVYVYMATRTTMTGRRMRKKE	5350
510	--LL-----V-----LK	514
5351	KCLLRQSRRLREDADVPEPTFK	5372

Search completed: January 3, 2005, 15:28:12  
Job time : 94.6667 secs

This Page Blank (uspto)



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:55:11 ; Search time 102.333 Seconds  
(without alignments)  
2889.993 Million cell updates/sec

Title: US-09-987-701-4

Perfect score: 3649

Sequence: 1 MSISDEVNPLVRYLQESG.....LPHLVVFVPLVLIALLVLK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot\_02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3437	94.2	514	1 TBLR HUMAN	Q9bk7 homo sapien
2	3405	93.3	514	1 TBLR MOUSE	Q8bhj5 mus musculus
3	3376.2	92.5	522	2 Q6GPC6	Q6gpc6 xenopus lae
4	3367.5	92.3	519	2 Q7SZM9	Q7szm9 xenopus lae
5	3216.1	88.1	527	1 TBLX MOUSE	Q9qxe7 mus musculus
6	3154.2	86.4	526	1 TBLX HUMAN	Q60907 homo sapien
7	3085.8	84.6	522	1 TBLY HUMAN	Q9bq87 homo sapien
8	2756.7	75.5	511	2 Q7Q371	Q7q371 anopheles g
9	2753.9	75.5	700	1 EBI DROME	Q95xj9 drosophila
10	2599.6	71.2	7073	2 AAP49011	Aap49011 sars coro
11	2598.3	71.2	20925	2 Q8NUX1	Q8nux1 trichoderma
12	2596.5	71.2	5636	2 Q6RWM7	Q96xw7 homo sapien
13	2592.4	71.0	34350	2 Q6WZ42	Q8wz42 homo sapien
14	2590.7	71.0	7191	2 Q6XA09	Q6xa09 alternaria
15	2590.7	71.0	7191	2 Q6XA09	Q6xa09 alternaria
16	2590.6	71.0	7073	2 AAP82978	Aap82978 sars coro
17	2589.5	71.0	4599	1 LRLB MOUSE	Q9jli8 mus musculus
18	2589.3	71.0	7962	2 Q6RCX8	Q10465 homo sapien
19	2588.6	70.9	7073	1 R1AB CVHSA	P59641 h replicase
20	2588.6	70.9	7073	2 Q6JH48	Q6jh48 sars corona
21	2588.6	70.9	7073	2 Q6R7Y8	Q6r7y8 sars corona
22	2588.6	70.9	7073	2 Q6RCW7	Q6rcw7 sars corona
23	2588.6	70.9	7073	2 Q6RCX8	Q6rcx8 sars corona
24	2588.6	70.9	7073	2 Q6RCY9	Q6rcy9 sars corona
25	2588.6	70.9	7073	2 Q6RD00	Q6rd00 sars corona
26	2588.6	70.9	7073	2 Q6RD11	Q6rd11 sars corona
27	2588.6	70.9	7073	2 Q6RD22	Q6rd22 sars corona
28	2588.6	70.9	7073	2 Q6RD33	Q6rd33 sars corona
29	2588.6	70.9	7073	2 Q6RD44	Q6rd44 sars corona
30	2588.6	70.9	7073	2 Q6RD55	Q6rd55 sars corona
31	2588.6	70.9	7073	2 Q6RD66	Q6rd66 sars corona

#### RESULT 1

ID	TBLR HUMAN	STANDARD;	PRT;	514 AA.
AC	Q9BZK7; Q9H2I1; Q9H9A1;			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	01-OCT-2004 (Rel. 45, Last annotation update)			
DE	F-box-like/WD-repeat protein TBLR1 (Nuclear corepressor/corepressor/HDAC3 complex subunit TBLR1) (TBL1-related protein 1).			
DE	complex subunit TBLR1			
GN	Name=TBLR1; Synonyms=IRAL;			
OS	Homo sapiens (Human)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR COMPLEX WITH NCOR1; NCOR2; GPS2; TBL1X AND HDAC3.			
RP	MEDLINE=21930350; PubMed=11931768;			
RA	Zhang J., Kalkum M., Chait B.T., Roeder R.G.;			
RT	"The N-Cor-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2.";			
RL	Mol. Cell 9:611-623(2002).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	PubMed=11063877;			
RA	Zhang X., Dormady S.P., Basch R.S.;			
RT	"Identification of four human CDNAs that are differentially expressed by early hematopoietic progenitors.";			
RL	Exp. Hematol. 28:1286-1296(2000).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RP	PubMed=14702039; DOI=10.1038/ngl1285;			
RA	Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S., Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahata K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa K., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K., Tanai H., Kimata M., Watanabe M., Hirakawa S., Chiba Y., Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J., Kanehori K., Takahashi-Fujii A., Hara H., Tanase T.-O., Nomura Y., Togiya S., Komai F., Hara R., Takeuchi K., Arita M., Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S., Yoshioka Y., Matsunawa H., Ichihara T., Shiohara H., Sano S., Moriya S., Momiya H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,			

#### ALIGNMENTS

32	2588.6	70.9	7073	2	Q6TPE9	Q6tpe9 sars corona
33	2588.6	70.9	7073	2	Q6UZF1	Q6uzf1 sars corona
34	2588.6	70.9	7073	2	Q6UZF5	Q6uzf5 sars corona
35	2588.6	70.9	7073	2	Q6V586	Q6v586 sars corona
36	2588.6	70.9	7073	2	Q6VA80	Q6va80 sars corona
37	2588.6	70.9	7073	2	Q6VA91	Q6va91 sars corona
38	2588.6	70.9	7073	2	Q6VAA2	Q6vaa2 sars corona
39	2588.6	70.9	7073	2	AAR14802	Aar14802 sars coro
40	2588.6	70.9	7073	2	AAR14806	Aar14806 sars coro
41	2588.6	70.9	7073	2	AAR14810	Aar14810 sars coro
42	2588.6	70.9	7073	2	AAR91584	Aar91584 sars coro
43	2588.6	70.9	7073	2	AAS000002	Aas00002 sars coro
44	2588.6	70.9	7073	2	AAP94735	Aap94735 sars coro
45	2588.6	70.9	7073	2	AAP94746	Aap94746 sars coro

RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama H., Hata H., Watanabe M., Komatsu T.,  
RA Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RL Nat. Genet. 36:40-45 (2004).  
RN [4]  
RN SEQUENCE OF 1-125 FROM N.A.  
RP TISSUE=Adrenal cortex;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavani T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueudin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [5]  
RN COMPONENT OF THE N-COR COMPLEX WITH TBL1X; CORO2A AND HDAC3, AND  
RP HISTONE-BINDING.  
RX PubMed=12628926;  
RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,  
RA Wong J.;  
RT "Purification and functional characterization of the human N-Cor  
RT complex: the roles of HDAC3, TBL1 and TBL1X";  
RL EMBO J. 22:1336-1346 (2003).  
RN [6]  
RN FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.  
RX PubMed=14980219;  
RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;  
RT "A corepressor/coactivator exchange complex required for  
RT transcriptional activation by nuclear receptors and other regulated  
RT transcription factors";  
RL Cell 116:511-526 (2004).  
CC -|- FUNCTION: F-box-like protein involved in the recruitment of the  
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated  
CC transcription units. Plays an essential role in transcription  
CC activation mediated by nuclear receptors. Probably acts as  
CC integral component of the N-Cor corepressor complex that mediates  
CC the recruitment of the 19S proteasome complex, leading to the  
CC subsequent proteasomal degradation of N-Cor complex, thereby  
CC allowing cofactor exchange, and transcription activation.  
CC -|- SUBUNIT: Component of the N-Cor repressor complex, at least  
CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.  
CC Probable component of some E3 ubiquitin ligase complex. Interacts  
CC with histones H2B and H4.  
CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -|- TISSUE SPECIFICITY: Ubiquitous.  
CC -|- DOMAIN: The F-box-like domain is related to the F-box domain, and  
CC apparently displays the same function as component of ubiquitin E3  
CC ligase complexes (By similarity).  
CC -|- SIMILARITY: Belongs to the WD-repeat E3 family.  
CC -|- SIMILARITY: Contains 1 F-box-like domain.  
CC -|- SIMILARITY: Contains 1 Lish domain.  
CC -|- SIMILARITY: Contains 8 WD repeats.  
CC -|- CAUTION: Ref.4 sequence differs from that shown due to a potential  
CC poly-A sequence starting in position 125.

CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC -----  
DR EMBL; AF314544; AAK00301.1; -;  
DR EMBL; AF268193; AAG44736.1; -;  
DR EMBL; AK022956; BAB14331.1; -;  
DR EMBL; BC060320; AAH60320.1; ALT\_TERM.  
DR HSSP; P16649; 1ERJ.  
DR MIM; 608628; -;  
DR InterPro; IPR006594; Lish.  
DR InterPro; IPR011047; Quin\_alc\_DH\_like.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF04000; WD40; 8.  
DR PRINTS; PR00320; GPROTEINRPT.  
DR ProDom; PD000018; WD40; 3.  
DR SMART; SM00667; Lish; 1.  
DR SMART; SM00320; WD40; 8.  
DR PROSITE; PS00896; LISH; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 4.  
DR PROSITE; PS00082; WD\_REPEATS\_2; 6.  
DR PROSITE; PS0294; WD\_REPEATS\_REGION; 1.  
KW Activator; Nuclear protein; Repeat; Transcription regulation;  
KW Ub1 conjugation pathway; WD repeat.  
FT DOMAIN 4 36  
FT Lish.  
FT DOMAIN 41 86  
FT F-box-like.  
FT REPEAT 167 206  
FT WD 1.  
FT REPEAT 223 282  
FT WD 2.  
FT REPEAT 264 303  
FT WD 3.  
FT REPEAT 306 344  
FT WD 4.  
FT REPEAT 347 386  
FT WD 5.  
FT REPEAT 389 437  
FT WD 6.  
FT REPEAT 440 479  
FT WD 7.  
FT REPEAT 481 513  
FT WD 8.  
FT DOMAIN 108 118  
FT Poly-Ala.  
FT CONFLICT 31 31  
FT E -> K (in Ref. 2).  
FT CONFLICT 59 59  
FT Y -> H (in Ref. 3).  
FT CONFLICT 389 389  
FT A -> Q (in Ref. 2).  
SQ SEQUENCE 514 AA; 55595 MW; 0B556D2EE4BA796D CRC64;  
Query Match 94.2%; Score 3437; DB 1; Length 514;  
Best Local Similarity 90.4%; Pred. No. 5.7e-62;  
Matches 483; Conservative 6; Mismatches 5; Indels 40; Gaps 12;  
QY 1 MSISDEVNFLVRYLYQESGFSGHSAFTFGIKSHISOSNINGALVPPAALISIIQKQYV 60  
DB 1 MSISDEVNFLVRYLYQESGFSGHSAFTFGIESHISOSNINGALVPPAALISIIQKQYV 60  
QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQQAAAAAASQ 120  
DB 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQQQAAAAAASQ 120  
QY 121 QGSAXNGENTANGEANGAHTIANNHTDMMEDVDGVEIPNKAVLRGHESEVPICAWNPV 180  
DB 121 QGSAXNGENTANGEANGAHTIANNHTDMMEDVDGVEIPNKAVLRGHESEVPICAWNPV 180  
QY 181 SLLASGSGDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPNSKNDVTSLDWNSGTL 240  
DB 181 SLLASGSGDSTARINWLSNSTSGSTQVLVLRHCIREGGQDVPNSKNDVTSLDWNSGTL 240  
QY 241 ATGSYDGFARIWTKDGNLSTLQHKGPFLAKWNKKNFILSAGVDKTTIIMDAHTGEA 300  
DB 241 ATGSYDGFARIWTKDGNLSTLQHKGPFLAKWNKKNFILSAGVDKTTIIMDAHTGEA 300  
QY 301 KQOPPHSPALDQVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360  
DB 301 KQOPPHSPALDQVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360

QY 361 GNLASCSDDMTLKIMSKQDNCVHDLQOHNKEIYTIKWSPTGPGTNNPNANMLASASP 420  
 DB 361 GNLASCSDDMTLKIMSKQDNCVHDLQOHNKEIYTIKWSPTGPGTNNPNANMLASASP 420  
 QY 421 DSTVRLWDVDRGICHTLTKHOEPVYVSAFSPDGRVLAGSDPKCVHINTQ-----VCLH 476  
 DB 421 DSTVRLWDVDRGICHTLTKHOEPVYVSAFSPDGRVLAGSDPKCVHINTQGTALV--H 478  
 QY 477 -Y--LNG--QVLN-----LGR-----SICLYTLPHLHVLPVLIALLVL-K 514  
 DB 479 SYRGTGIEVCWNAAGDKVGASDGSVC-----V-L-D-----LRK 514

RESULT 2  
 TBLR MOUSE  
 ID TBLR MOUSE STANDARD; PRT; 514 AA.  
 AC Q8BRJ5; Q8CEG4; Q8VEG3; Q9EQD4;  
 DT 05-JUL-2004 (Rel. 44, Created)  
 DT 05-JUL-2004 (Rel. 44, Last sequence update)  
 DE F-box-like/WD-repeat protein TBLR1 (Nuclear receptor corepressor/HDAC3  
 DE complex subunit TBLR1) (TBL1-related protein 1).  
 GN Name:Tblr1; Synonyms:Iral;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c;  
 RA Zhang X., Dormady S., Basch R.;  
 RT "Identification of four human cDNAs that are differentially expressed  
 RT by early hematopoietic progenitors";  
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Testis;  
 RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
 RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
 RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
 RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
 RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
 RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
 RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,  
 RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
 RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
 RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
 RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
 RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
 RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
 RA Nagashima T., Numata K., Okido T., Pavan W.J., Perteza G., Pesole G.,  
 RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
 RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
 RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,  
 RA Sulkana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
 RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
 RA Wilming L.G., Wyszynski B., Yanagisawa M., Yang I., Yang L.,  
 RA Yuan Z., Zavalan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
 RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
 RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
 RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
 RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,  
 RA Yasunishi A., Yoshino K., Waterston R., Lander E.S., Rogers J.,  
 RA Birney E., Hayashizaki Y.;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs";  
 RL Nature 420:563-573(2002).  
 RN [3]  
 RP SEQUENCE OF 314-514 FROM N.A.  
 RC TISSUE=Breast tumor;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 CC -!- FUNCTION: F-box-like protein involved in the recruitment of the  
 CC ubiquitin/19S proteasome complex to nuclear receptor-regulated  
 CC transcription units. Plays an essential role in transcription  
 CC activation mediated by nuclear receptors. Probably acts as  
 CC integral component of the N-CoR corepressor complex that mediates  
 CC the recruitment of the 19S proteasome complex, leading to the  
 CC subsequent proteosomal degradation of N-CoR complex, thereby  
 CC allowing cofactor exchange, and transcription activation (By  
 CC similarity).  
 CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least  
 CC composed of NCOR1, NCOR2, HDAC3, TBLR1, TBLR2, and GPS2.  
 CC Probable component of some E3 ubiquitin ligase complex. Interacts  
 CC with histones H2B and H4 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and  
 CC apparently displays the same function as component of ubiquitin E3  
 CC ligase complexes (By similarity).  
 CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.  
 CC -!- SIMILARITY: Contains 1 F-box-like domain.  
 CC -!- SIMILARITY: Contains 8 WD repeats.  
 CC -!- SIMILARITY: Contains 8 WD repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF268195; AAC44738.1; -;  
 CC EMBL; AK029595; BAC26526.1; -;  
 CC EMBL; AK033347; BAC28241.1; -;  
 CC EMBL; AK036064; BAC29294.1; -;  
 CC EMBL; BC018512; AAH18512.1; -;  
 CC HSSP; P16649; 1ERJ.  
 CC InterPro; IPR006594; Lish.  
 CC InterPro; IPR011047; Quin\_ald\_DH\_like.  
 CC InterPro; IPR001680; WD40.  
 CC Pfam; PF04000; WD40; 8.  
 CC PRINTS; PR00320; GPROTEINBRPT.  
 CC ProDom; PD000018; WD40; 3.  
 CC SMART; SM00667; Lish; 1.  
 CC SMART; SM00320; WD40; 8.  
 CC PROSITE; PS50896; LISH; 1.  
 CC PROSITE; PS00678; WD\_REPEATS\_1; 4.  
 CC PROSITE; PS50082; WD\_REPEATS\_2; 6.  
 CC PROSITE; PS50294; WD\_REPEATS\_REGION; 1.  
 CC KW Activator; Nuclear protein; Repeat; Transcription regulation;  
 CC Ubl conjugation pathway; WD repeat.  
 CC FT DOMAIN 4 36 Lish.  
 CC FT DOMAIN 41 86 F-box-like.  
 CC FT REPEAT 167 206 WD 1.  
 CC FT REPEAT 223 262 WD 2.  
 CC FT REPEAT 264 303 WD 3.

```

FT REPEAT 306 344 WD 4.
FT REPEAT 347 386 WD 5.
FT REPEAT 389 437 WD 6.
FT REPEAT 440 479 WD 7.
FT REPEAT 481 513 WD 8.
FT DOMAIN 108 117 Poly-Ala.
FT CONFLICT 185 185 A -> V (in Ref. 1).
FT CONFLICT 374 374 K -> R (in Ref. 2; BAC29294).
FT CONFLICT 480 480 Y -> C (in Ref. 3).
SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;

Query Match 93.3%; Score 3405; DB 1; Length 514;
Best Local Similarity 89.5%; Pred. No. 2.8e-61;
Matches 478; Conservative 7; Mismatches 9; Indels 40; Gaps 12;

QY 1 MSISDEVNFLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNFLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAAASQ 120
Db 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAAATNQ 120

QY 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFICAWNPV 180
Db 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVFICAWNPV 180

QY 181 SDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240
Db 181 SDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240

QY 241 ATGSDGFAIRWTKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTIIWDAHTGEA 300
Db 241 ATGSDGFAIRWTKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTIIWDAHTGEA 300

QY 301 KQGFPHSPALDVDMQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGTHTNEVNAIKWDPT 360
Db 301 KQGFPHSPALDVDMQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGTHTNEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKIWSMKQDCNVHDLQHNKEIYIKSPGPGTNNPNANIMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDCNVHDLQHNKEIYIKSPGPGTNNPNANIMLASAF 420

QY 421 DSTVRLWDVDRGICHTLTKHQSPVSVAFSPDGRVYLASGPKCVHIWNTQ----VCLH 476
Db 421 DSTVRLWDVDRGICHTLTKHQSPVSVAFSPDGRVYLASGPKCVHIWNTQGTALV--H 478

QY 477 -Y--LNG--QVLN-----LGR-----SICLYTLPHLHVVIPLVLIELVL-K 514
Db 479 SYRGTGIFVVCWNAAGDKVGASGDSVC-----V-L-D---LRK 514

```

## RESULT 3

```

Q6GPC6 ID Q6GPC6 PRELIMINARY; PRT; 522 AA.
AC Q6GPC6;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh P.,

```

```

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.R., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Frange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyzanski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RL and mouse cDNA sequences.";
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391(2002).
[3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 8 WD repeats.
DR EMBL: BC073215; AAH73215.1; -.
DR InterPro: IPR006594; Lish; 1.
DR InterPro: IPR011047; Quin_alc_DH_like.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 3.
DR SMART: SM00667; Lish; 1.
DR SMART: SM00320; WD40; 8.
DR PROSITE: PS00896; Lish; 1.
DR PROSITE: PS00678; WD_REPEATS_1; 4.
DR PROSITE: PS00882; WD_REPEATS_2; 6.
DR PROSITE: PS0294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 522 AA; 56284 MW; 4FB0F58F503B461F CRC64;

```

```

Query Match 92.5%; Score 3376.2; DB 2; Length 522;
Best Local Similarity 87.5%; Pred. No. 1.2e-60;
Matches 474; Conservative 9; Mismatches 11; Indels 48; Gaps 15;

```

```

QY 1 MSISDEVNFLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNFLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYV 60

QY 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQO-----AAAAA 116
Db 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQQOAAAAA 120

QY 117 AA-S---QQGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEV 172
Db 121 AATSANNQPPAKNGESTANGEANGHALANNHTDMMEVDGVEIPPNKAVVLRGHESEV 180

QY 173 FICAWNPVSDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLD 232
Db 181 FICAWNPVSDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLD 240

QY 233 WNSEGTLATGSDGFAIRWTKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTII 292
Db 241 WNSEGTLATGSDGFAIRWTKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTII 300

QY 293 WDAHTGEAKQQPPFHSAPALDVDMQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGTHTNEV 352

```

```

Db 301 WDAHTGAQKQPPHSPALDWDQSNNTFASCTDMCIHVCKLQGDPRPKTFQGHTEV 360
Qy 353 NAIKWDPTGNLLASCSDMTLKIWSMKQDCVHDLQOHNKIYTIKWSPTGPTGNPNAN 412
Db 361 NAIKWDPTGNLLASCSDMTLKIWSMKHDTCVHDLQAHNKIYTIKWSPTGPTGNPNAN 420
Qy 413 LMLASASFDSTVRLWDVDRGICHTLTKHQPVPYVAFSPDGRYLASGSPDKCVHIWNTQ 472
Db 421 LMLASASFDSTVRLWDVDRGICHTLTKHQPVPYVAFSPDGRYLASGSPDKCVHIWNTQ 480
Qy 473 -----VCLH-Y--LNG--QVLLN-----LGR-----SICLYTLPHLWVPLVALIELLVL 513
Db 481 TGAIV--HSYRTGGIFECVNAAGDKVGASASDGSVC-----V-L-D---L 520
Qy 514 -K 514
Db 521 RK 522

RESULT 4
ID Q7SZM9 PRELIMINARY; PRT; 519 AA.
AC Q7SZM9;
DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Nuclear receptor co-repressor complex subunit TBLR1.
OS Xenopus laevis (African clawed frog)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=12794076;
RA Tomita A., Buchholz D.R., Obata K., Shi Y.B.;
RT "Fusion Protein of Retinoic Acid Receptor [alpha] with Promyelocytic
RT Leukemia Protein or Promyelocytic Leukemia Zinc Finger Protein
RT Recruits N-CoR-TBLR1 Corepressor Complex to Repress Transcription in
RT Vivo."
RL J. Biol. Chem. 278:30788-30795 (2003).
CC -I- SIMILARITY: Contains 8 WD repeats.
DR EMBL; AY225088; AAP20646.1; -.
DR GO; GO:004872; F:receptor activity; IEA.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS00882; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Receptor; Repeat; WD repeat.
SQ SEQUENCE 519 AA; 56043 MW; 5E998EDC8C92296 CRC64;

Query Match 92.3%; Score 3367.5; DB 2; Length 519;
Best Local Similarity 87.8%; Pred. No. 1.9e-60;
Matches 472; Conservative 11; Mismatches 11; Indels 45; Gaps 14;

Qy 1 MSISDEVNPLVRYLQESGFSHTFFGKSHISQSNINGALVPPAALISIIQKGLQV 60
Db 1 MSISDEVNPLVRYLQESGFSHTFFGKSHISQSNINGALVPPAALISIIQKGLQV 60
Qy 61 EAFVSINEDGTLFDGRPIESLSLIDAVPDVQTRQAYRDKLAQQQ--AAAAA 118
Db 61 EAFVSINEDGTLFDGRPIESLSLIDAVPDVQTRQAYRDKLAQQQTRAAAAA 120
Qy 119 S---QQGSAKNGTANGENGAAHTTANNHTDMMEDVDGVEIPPNKAVVLRGHSESVFC 175
Db 121 TPNQOPPAKNGTANGENGAAHTTANNHTDMMEDVDGVEIPSPKAVVLRGHSESVFC 180

```

## RESULT 5

TBLX\_MOUSE STANDARD; PRT; 527 AA.

```

ID TBLX_MOUSE STANDARD; PRT; 527 AA.
AC Q9QXE7; Q8BMM0; Q8BYQ4; Q8COAL;
DT 16-OCT-2001 (Rel. 40, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/wd-repeat protein TBLX (Transducin beta-like 1X protein).
GN Name=Tblx; Synonyms=Tbll;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RC TISSUE=Hypothalamus, Medulla oblongata, and Pituitary;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Kongsaga A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltaga K., Marchionni L., Mckenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi F., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takemura Y., Taylor M.S., Teasdale R.D., Tomita K.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wymshaw-Boris A., Yanagisawa M., Yang L., Yang L.,
RA Yuan Z., Zavalon M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs."
RL Nature 420:563-573 (2002).
RN [2]

```



Db 523 -L-D---LRK 527

RESULT 6

TBLX\_HUMAN

ID TBLX\_HUMAN STANDARD; PRT; 526 AA.

AC O60907; Q86UY2; 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)

DE F-box-like/WD-repeat protein TBL1X (Transducin beta-like IX protein)

DE (Transducin-beta-like 1, X-linked) (SMP55).

GN Name=TBL1X; Synonyms=TBL1;

OS Homo sapiens (Human)

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DISEASE.

RX MEDLINE=99264241; PubMed=10330347;

RA Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,

RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.,

RT "X-linked late-onset sensorineural deafness caused by a deletion

RT involving OAI and a novel gene containing WD-40 repeats.";

RL Am. J. Hum. Genet. 64:1604-1616(1999).

RN [2]

RP SEQUENCE FROM N.A.

RX TISSUE=Lymph, and Pancreas;

RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klauener R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

RA Scapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,

RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,

RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,

RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,

RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,

RA Fahey J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,

RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,

RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,

RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,

RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,

RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;

RT "Generation and initial analysis of more than 15,000 full-length human

RT and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [3]

RP IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR

RP COMPLEX WITH NCOR2 AND HDAC3.

RX PubMed=10809664;

RA Guenther M.G., Lane W.S., Fischle W., Verdin E., Lazar M.A.,

RA Shiekhattar R.;

RT "A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-

RT repeat protein linked to deafness.";

RL Genes Dev. 14:1048-1057(2000).

RN [4]

RP COMPONENT OF THE N-COR COMPLEX WITH NCOR2 AND HDAC3.

RX PubMed=10944117;

RA Li J., Wang J., Wang J., Nawaz Z., Liu J.M., Qin J., Wong J.;

RT "Both corepressor proteins SMRT and N-COR exist in large protein

RT complexes containing HDAC3.";

RL EMBO J. 19:4342-4350(2000).

RN [5]

RP SUBUNIT OF A COMPLEX WITH UBE2D1, CACYBP, SIAH1 AND APC.

RX PubMed=11389839;

RA Matsuzawa S.-I., Reed J.C.;

RT "Siah-1, SIP, and Ebi collaborate in a novel pathway for beta-catenin

RT degradation linked to p53 responses.";

RL Mol. Cell 7:915-926(2001).

RN [6]

RP COMPONENT OF THE N-COR COMPLEX WITH NCOR1, NCOR2, GPS2, TBL1R AND

RP HDAC3.

RX MEDLINE=21930350; PubMed=11931768;

RA Zhang J., Kalkum M., Chait B.T., Roeder R.G.;

RT "The N-COR-HDAC3 nuclear receptor corepressor complex inhibits the JNK

RT pathway through the integral subunit GPS2.";

RL Mol. Cell 9:611-623(2002).

RN [7]

RP COMPONENT OF THE N-COR COMPLEX WITH TBL1R, CORO2A AND HDAC3, AND

RP HISTONE-BINDING.

RX PubMed=12628926;

RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,

RA Wong J.;

RT "Purification and functional characterization of the human N-COR

RT complex: the roles of HDAC3, TBL1 and TBL1R.";

RL EMBO J. 22:1336-1346(2003).

RN [8]

RP FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.

RX PubMed=14980219;

RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;

RT "A corepressor/coactivator exchange complex required for

RT transcriptional activation by nuclear receptors and other regulated

RT transcription factors.";

RL Cell 116:511-526(2004).

CC -!- FUNCTION: F-box-like protein involved in the recruitment of the

CC ubiquitin/19S proteasome complex to nuclear receptor-regulated

CC transcription units. Plays an essential role in transcription

CC activation mediated by nuclear receptors. Probably acts as

CC integral component of corepressor complexes that mediates the

CC recruitment of the 19S proteasome complex, leading to the

CC subsequent proteasomal degradation of transcription repressor

CC complexes, thereby allowing cofactor exchange.

CC -!- SUBUNIT: Component of the N-Cor repressor complex, at least

CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.

CC Component of a E3 ubiquitin ligase complex containing UBE2D1,

CC SIAH1, CACYBP/SIP, SKP1A, APC and TBL1X. Probably part of other

CC corepressor complexes, that do not contain NCOR1 and NCOR2.

CC Interacts with histones H2B, H3a and H4.

CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).

CC -!- TISSUE SPECIFICITY: Ubiquitous.

CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and

CC apparently displays the same function as component of ubiquitin E3

CC ligase complexes (By similarity).

CC -!- DISEASE: Defects in TBL1X may be a cause of late-onset

CC sensorineural deafness (OASD) (MIM:300650). OASD is an X-linked

CC recessive disorder characterized by ocular albinism and

CC progressive sensorineural hearing loss in the fourth and fifth

CC decades of life. OASD may be caused by deletion of both GPR143/OAI

CC and TBL1X adjacent genes; TBL1X defects possibly causing the

CC hearing phenotype.

CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.

CC -!- SIMILARITY: Contains 1 F-box-like domain.

CC -!- SIMILARITY: Contains 1 Lish domain.

CC -!- SIMILARITY: Contains 8 WD repeats.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).

CC -----

DR EMBL; Y12781; CAA73319.1; ALT\_INIT.

DR EMBL; BC032708; AAH32708.1; -.

DR EMBL; BC052304; AAH52304.1; ALT\_INIT.

DR HSSP; P16649; 1ERJ.

DR Genew; HGNC:11585; TBL1X.

DR MIM; 300196; -.

DR MIM; 300650; -.

DR GO; GO:0007605; P:perception of sound; TAS.

DR GO; GO:0007165; P:signal transduction; TAS.

DR GO; GO:0007601; P:visual perception; TAS.



```

DR InterPro: IPR006594; Lish.
DR InterPro: IPR011047; Quin_alc_DH_like.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 3.
DR SMART: SM00667; Lish; 1.
DR SMART: SM00320; WD40; 8.
DR PROSITE: PS00896; Lish; 1.
DR PROSITE: PS00678; WD REPEATS 1; 4.
DR PROSITE: PS00082; WD REPEATS 2; 6.
DR PROSITE: PS00294; WD REPEATS REGION; 1.
KW Activator; Deafness; Nuclear protein; Repeat;
KW Transcription regulation; Ub1 conjugation pathway; WD repeat.
FT DOMAIN 4 36 Lish.
FT REPEAT 179 218 WD 1.
FT REPEAT 235 274 WD 2.
FT REPEAT 276 315 WD 3.
FT REPEAT 318 358 WD 4.
FT REPEAT 359 398 WD 5.
FT REPEAT 401 449 WD 6.
FT REPEAT 452 491 WD 7.
FT REPEAT 493 525 WD 8.
FT DOMAIN 108 124 Poly-Ala.
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F659 CRC64;

Query Match      86.4%; Score 3154.2; DB 1; Length 526;
Best Local Similarity 78.3%; Pred. No. 8e-56;
Matches 430; Conservative 41; Mismatches 20; Indels 58; Gaps 22;

Qy 1 MSISDEVNVLVRYLQESGFSHSATFGIKSHISOSNINGALVPAALISIIKGLQYV 60
Db 1 MSISDEVNVLVRYLQESGFSHSATFGIKSHISOSNINGALVPAALISIIKGLQYV 60

Qy 61 EAEVSINEDGTLDGRIEISLIDAMPDVVOTROQAVRDKLAQQ--AAAAAAA-AA 117
Db 61 EAEVSINEDGTLDGRIEISLIDAMPDVVOTROQAFREKLAQQASAAAAATAA 120

Qy 118 ASQ-----QG-S-----AKNGENTANGENGHAHTIANHTDM-MEVDGDVEIPNKAVALR 166
Db 121 ATAATTSAGVSHQNFPSKNREATVNGEENRAHSV-NNHA-KPMEIDGVEIPSSKATVLR 178

Qy 167 GHESEVFIQAWNPVSDLLASGSGDSTARIWNLSENSTSGTQVLVLRHCTREGGQDVPSNK 226
Db 179 GHESEVFIQAWNPVSDLLASGSGDSTARIWNLSENNGSGTQVLVLRHCTREGGHDVPSNK 238

Qy 227 DVTSLDMNSGTLATGSDYDGFARIWTKDGNLASTLQGHKGPFPALKWKKGNFILLSAGV 286
Db 239 DVTSLDMNTGTLATGSDYDGFARIWTEGDNLASTLQGHKGPFPALKWKNKGNVILLSAGV 298

Qy 287 DKTITLWDAHTGAKQFPFHSAPALDVQSN--TFASCTDMCHVCKLGQDRIPKTF 345
Db 299 DKTITLWDAHTGAKQFPFHSAPALDVQ--NNTTFASCTDMCHVCKLGCDRFPKTF 357

Qy 346 QGHTNEVNAIKWPTGNLLASCDDMTLTKIWSMKQDNCVHDLQOHKNEIYTIKWSPTGPG 405
Db 358 QGHTNEVNAIKWPTGNLLASCDDMTLTKIWSMKQEVCHDLQAHKNEIYTIKWSPTGPA 417

Qy 406 TNNPNANMLASFSFSTVRLWVDVRCIGTHTLTKHQEPVYSVAFSPDGRYLASGFDKC 465
Db 418 TSNPNSNIMLASFSFSTVRLWVDIERGVCTHTLTKHQEPVYSVAFSPDGRYLASGFDKC 477

Qy 466 VHTWNTQ-----VCLH-Y--LNG--QVLLN-----LGR-----SICLYTLPHLWVPLVA 506
Db 478 VHTWNTQSNLV--HSYRGTTGGTFFECWNRAGDKVGASADGSGVC-----V- 521

Qy 507 LIELLVL-K 514
Db 522 L-D---LRK 526

RESULT 7

```

```

TBLY_HUMAN
ID TBLY_HUMAN STANDARD; PRT; 522 AA.
AC Q9BQ57;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein TBLY (Transducin beta-like 1V protein)
DE (Transducin-beta-like 1, Y-linked).
GN Name=TBLY; Synonyms=TBLL;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., AND TISSUE SPECIFICITY
RX MEDLINE=22699787; PubMed=12815422; DOI=10.1038/nature01722;
RA Skaletsky H., Kuroda-Kawaguchi T., Minx P.J., Cordum H.S.,
RA Hillier L.W., Brown L.G., Repping S., Pyntikova T., Ali J., Bieri T.,
RA Chinwalla A., Delehaanty A., Delehaanty K., Du H., Fewell G.,
RA Fulton L., Fulton R., Graves T.A., Hou S.-F., Latreille P.,
RA Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Naeh W.,
RA Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K.,
RA Schultz B., Strong C., Tin-Wollam A., Yang S.-P., Waterston R.H.,
RA Wilson R.K., Rozen S., Page D.C.;
RT "The male-specific region of the human Y chromosome is a mosaic of
RT discrete sequence classes.";
RL Nature 423:825-837(2003).
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated
CC transcription units. Plays an essential role in transcription
CC activation mediated by nuclear receptors. Probably acts as
CC integral component of corepressor complexes that mediates the
CC recruitment of the 19S proteasome complex, leading to the
CC subsequent proteosomal degradation of transcription repressor
CC complexes, thereby allowing cofactor exchange (By similarity).
CC -!- SUBUNIT: Probable component of the N-CoR repressor complex and
CC some E3 ubiquitin ligase complex (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- TISSUE SPECIFICITY: Fetal brain and prostate.
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and
CC apparently displays the same function as component of ubiquitin E3
CC ligase complexes (By similarity).
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.
CC -!- SIMILARITY: Contains 1 F-box-like domain.
CC -!- SIMILARITY: Contains 1 Lish domain.
CC -!- SIMILARITY: Contains 8 WD repeats.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF332220; AAK13472.1; -.
DR EMBL: AF332221; AAK13473.1; -.
DR EMBL: AF332222; AAK13474.1; -.
DR HSP: P16649; IERJ.
DR Genew; HGNC:18502; TBLY.
DR MIM; 400033; -.
DR InterPro: IPR006594; Lish.
DR InterPro: IPR001680; WD40.
DR InterPro: IPR011046; WD40_like.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 3.
DR SMART: SM00667; Lish; 1.
DR SMART: SM00320; WD40; 8.
DR PROSITE: PS00896; Lish; 1.
DR PROSITE: PS00678; WD REPEATS 1; 4.
DR PROSITE: PS00294; WD REPEATS 2; 6.
DR PROSITE: PS00294; WD REPEATS REGION; 1.

```



KW Activator; Nuclear protein; Repeat; Transcription regulation;  
KW Ubi conjugation pathway; WD repeat.  
FT DOMAIN 4 36 Lish.  
FT DOMAIN 41 86 F-box-like.  
FT REPEAT 177 216 WD 1.  
FT REPEAT 233 272 WD 2.  
FT REPEAT 274 313 WD 3.  
FT REPEAT 316 354 WD 4.  
FT REPEAT 357 396 WD 5.  
FT REPEAT 399 447 WD 6.  
FT REPEAT 450 489 WD 7.  
FT REPEAT 491 521 WD 8.  
SQ SEQUENCE 522 AA; 5668 MW; 4E020216422442D8 CRC64;  
  
Query Match 84.6%; Score 3085.8; DB 1; Length 522;  
Best Local Similarity 78.0%; Pred. No. 2.4e-54;  
Matches 423; Conservative 37; Mismatches 30; Indels 52; Gaps 22;  
  
QY 1 MSISDEVNPLVRYLQESGFSHSAFTFGIKSHISOSNINGALVPPAALISIIQKGLQYV 60  
DB 1 MSISDEVNPLVRYLQESGFSHSAFTFGIESHISOSNINGTLVPPSALISIIQKGLQYV 60  
  
QY 61 EAEVSINEDGTLPDGRPIESLSLIDAVMPDVVOTROQAYRDKLAQQA-AAA--A-A-A-114  
DB 61 EAEISINKDGVDFSPRIESLSLIVAVIPDVQMROQAFGEKLTQQQASAAATEASAMAK 120  
  
QY 115 AA-----AA-SQGSACKNGENTANGEAHTIANNHTDM-MEVDGVEIPPNKAVVLRGH 168  
DB 121 AATWTFAAISQNPKNRATVNGEAGAEI-NNHS-KPMEIDGVEIPPNKATVLRGH 178  
  
QY 169 ESEVFICAWNPVSDLLASGSGDSTARIMNLSNSTSGTQLVLRHCIRREGQDVPSNKDV 228  
DB 179 ESEVFICAWNPVSDLLASGSGDSTARIMNLSNSNGSTQLVLRHCIRREGHDVPSNKDV 238  
  
QY 229 TSLDWNSEGTILATGSDGFARITWKDGNLSTLGHQKGPFPALKKWKNGNFIISAGVDK 288  
DB 239 TSLDWNSEGTLLAMGSDGFARITWENGSLASTLGHQKGPFPALKKWKNGNYLSAGVDK 298  
  
QY 289 TTIINDAHTGEAKQPPFHSAPALPDVDWQSN--TPASCSTDMCIHVCKLGQRPKTG 347  
DB 299 TTIINDAHTGEAKQPPFHSAPALPDVDWQ--NNMTFASCSTDMCIHVCRGCDHPVKTF 357  
  
QY 348 HTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGFTN 407  
DB 358 HTNEVNAIKWDPSGMLLASCSDDMTLKIWSMKQDACVHDLQHSKEIYTIKWSPTGPTS 417  
  
QY 408 NPVNAIMLASAFDSTVRLWDVDRGICHTLTKHQBEPVYSVAFSPDGRYLASGSPDKVH 467  
DB 418 NPNSSIMLASAFDSTVRLWDVEQGVCTHTLMKHQBEPVYSVAFSPDGKYLASGSPDKYVH 477  
  
QY 468 IWNTO-----VCLH-Y--LNG--QVLN-----LGR-----SICLYTLPHLVVPIPLVALI 508  
DB 478 IWNTOGSLV--HSYQGTGGIPEVCNARGDKVGASASDGSVC-----V-L--520  
  
QY 509 EL 510  
DB 521 DL 522  
  
RESULT 8  
QY0371 PRELIMINARY; PRT; 511 AA.  
AC QY0371;  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE AgCp11563 (Fragment).  
GN Name=agc45469; ORFNames=ENSANG00000019755;  
OS Anopheles gambiae str. PREST.  
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;  
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
OX NCBI\_TaxID=180454;  
RN [1]

RP SEQUENCE FROM N.A.  
RC STRAIN=PEST;  
RA Anopheles Genome Sequencing Consortium;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
CC -!- SIMILARITY: Contains 8 WD repeats.  
CC -!- CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
DR EMBL; AAA01008964; EAA12470.1; --  
DR InterPro; IPR006594; Lish.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF04000; WD40; 8.  
DR PRINTS; PR00320; GPROTEINBRPT.  
DR ProDom; PD000018; WD40; 3.  
DR PROSITE; PS00896; LISH; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 3.  
DR PROSITE; PS00882; WD\_REPEATS\_2; 6.  
DR PROSITE; PS00294; WD\_REPEATS\_REGION; 1.  
KW Repeat; WD repeat.  
FT NON TER 1  
SQ SEQUENCE 511 AA; 55631 MW; 7C4C2670316A8580 CRC64;  
  
Query Match 75.5%; Score 2756.7; DB 2; Length 511;  
Best Local Similarity 68.4%; Pred. No. 3.1e-47;  
Matches 379; Conservative 49; Mismatches 33; Indels 93; Gaps 39;  
  
QY 1 MSISDEVNPLVRYLQESGFSHSAFTFGIKSHISOSNINGALVPPAALISIIQKGLQYV 60  
DB 8 MFSSDEVNPLVRYLQESGFSHSAFTFGIESHISOSNINGALVPPAALISIIQKGLQYV 67  
  
QY 61 EAEVSINEDGT---LFDGRPIESLSLIDAVMPDVVOTRO--QAYRDKLAQ--QAAAAA 114  
DB 68 EAEISIGEDGTQRL--G---ESLSLIDAVMEVAVNRQNMQ-----NQKQ----- 109  
  
QY 115 AAAASQGSACKNGENTANG--BENGHAH-----TIANNHT-----DMVEVDGVEIPPNKA 162  
DB 110 -ATKTEE---P--E-T-NCTTEE-PAPPPPTTTTATT-TPAPVETMEVDQSIETPASKA 159  
  
QY 163 VVLRGHESEVFICAWNPVS-DLLASGSGDSTARIMNLSNSTSGTQLVLRHCIRREGQD 221  
DB 160 TVLRGHESEVFICAWNP-STDLLASGSGDSTARIMNLSNSTSGTQLVLRHCIRREGQD 217  
  
QY 222 VPSNKKDVTSLDWNSEGTILATGSDGFARIW--TKDGNLSTLGHQKGPFPALKKWKNGNF 280  
DB 218 VPSNKKDVTSLDWNCDGTLATGSDGFARIWKT-DGLLASTLGHQKGPFPALKKWKNGNY 276  
  
QY 281 ILSAGVDKTTIINDAHTGE--AKQPPFHSAPALPDVDWQSNNTFASCSTDMCIHVCKLGQD 339  
DB 277 ILSAGVDKTTIINDAHTGCT--QQFSFHSAPALPDVDWQSNQSFASCSTDMCIHVCKLGVD 335  
  
QY 340 RIKTFOGHTNEVNAIKWDPTGNLLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYTIKW 399  
DB 336 KPIKFOGHTNEVNAIKWDPPQGLLASCSDDMTLKIWSMKQDTCVHDLQHSKEIYTIKW 395  
  
QY 400 SPTGPTGNPNANMLASAFDSTVRLWDVDRGICHTLTKHQBEPVYSVAFSPDGRYLAS 459  
DB 396 SPTGPTGNPNANMLASAFDSTVRLWDVDRGICHTLTKHQBEPVYSVAFSPDGRYLAS 455  
  
QY 460 GSFDKCVHIWNTQ-----VCLH-YLANGQVLNLSGRSI---CLY-----TLPHLV----- 500  
DB 456 GSFDKCVHIWNTSQGLV--HSY-KG-----TG-GIFEVC-WNSRGS-K---VGASASDG 501  
  
QY 501 -VPLVALIEL--L 511  
DB 502 SV--FV-L-DLRK 511  
  
RESULT 9  
EBI\_DROME  
ID EBI\_DROME STANDARD; PRT; 700 AA.  
AC Q95R39; Q9XZK1;  
DT 05-JUL-2004 (Rel. 44, Created)  
DT 05-JUL-2004 (Rel. 44, Last sequence update)

DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE F-box-like/WD-repeat protein ebi.  
 GN Name-ebi; ORFNames-CG4063;  
 OS Drosophila melanogaster. [Fruit fly].  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OC NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE  
 RP SPECIFICITY, AND MUTANT EBI4.  
 RX MEDLINE=99234084; PubMed=10215623;  
 RA Dong X., Tsuda L., Zavitz K.H., Lin M., Li S., Carthew R.W.,  
 RA Zipursky S.L.;  
 RT "ebi regulates epidermal growth factor receptor signaling pathways in  
 RT Drosophila.";  
 RL Genes Dev. 13:954-965 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=Berkely;  
 RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.B., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foaier C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaisai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milehina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RP STRAIN=Berkely; TISSUE=Embryo;  
 RX MEDLINE=22426066; PubMed=12537569;  
 RA Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarini H., Krommiller B., Pacieb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A Drosophila full-length cDNA resource.";  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).  
 RN [4]  
 RP FUNCTION, INTERACTIONS WITH PHYL AND SINA, AND MUTANTS EBI1C1 AND  
 RP EBI1C3.  
 RX PubMed=11032805;  
 RA Boulton S.J., Brook A., Staehling-Hampton K., Heitzler P., Dyson N.;  
 RT "A role for Ebi in neuronal cell cycle control.";  
 RL EMBO J. 19:5376-5386 (2000).  
 RN [5]  
 RP FUNCTION, AND INTERACTIONS WITH SNO; SMR AND SU(H).  
 RP PubMed=12230979;  
 RA Tsuda L., Nagaraj R., Zipursky S.L., Banerjee U.;  
 RT "An EGFR/Ebi/Sno pathway promotes delta expression by inactivating  
 RT Su(H)/SMRTER repression during inductive notch signaling.";  
 RL Cell 110:625-637 (2002).  
 CC -!- FUNCTION: F-box-like component of E3 ubiquitin ligase complexes;  
 CC involved in R7 photoreceptor cell differentiation, cone cell  
 CC development and neuronal cell cycle control. E3 ubiquitin ligase  
 CC complexes mediate ubiquitination and subsequent proteasomal  
 CC degradation of target proteins. Required for specification of R7  
 CC photoreceptor cell fate in the eye by participating to the  
 CC ubiquitination and subsequent proteasomal degradation of Tramtrack  
 CC (ttk), a general inhibitor of photoreceptor differentiation.  
 CC Required to block the S phase entry in the peripheral nervous  
 CC system and central nervous system in a process that does not  
 CC involve the degradation of ttk. Involved in cone cell development  
 CC by preventing the transcriptional repression mediated by Su(H) on  
 CC D1, probably by participating to a E3 complex that contains sno  
 CC degradation of some component of the Su(H) repressor complex.  
 CC -!- SUBUNIT: Component of some E3 complex at least composed of sina,  
 CC ebi and phyl, required for the degradation of ttk. Probably forms  
 CC a E3 complex with sno, required for the degradation of some  
 CC component of the Su(H) repressor complex. Interacts with sno and  
 CC Su(H) and Smr.  
 CC -!- SUBCELLULAR LOCATION: Nuclear.  
 CC -!- TISSUE SPECIFICITY: Widely expressed both in embryos and larvae.  
 CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and  
 CC apparently displays the same function as component of ubiquitin E3  
 CC ligase complexes.  
 CC -!- MISCELLANEOUS: 'Ebi' means 'shrimp' in Japanese.  
 CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.  
 CC -!- SIMILARITY: Contains 1 F-box-like domain.  
 CC -!- SIMILARITY: Contains 1 Lish domain.  
 CC -!- SIMILARITY: Contains 8 WD repeats.  
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to a  
 CC frameshift in position 484.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AF146345; AAD35017.1; --  
 DR EMBL: AB003589; AAF51501.1; --  
 DR EMBL: AY061326; AAL28874.1; ALT\_FRAME.  
 DR HSSP: P16649; 1ERJ.  
 DR FlyBase: FBgn0023444; ebi.  
 DR GO: GO:0000074; P:regulation of cell cycle; IMP.  
 DR InterPro: IPR006594; Lish.  
 DR InterPro: IPR001680; WD40.  
 DR Pfam: PF00400; WD40; 3.  
 DR PRINTS: PR00320; GPROTEINRPT.  
 DR ProDom: PD000018; WD40; 3.  
 DR SMART: SM00667; Lish; 1.  
 DR SMART: SM00320; WD40; 8.  
 DR PROSITE: PS00896; Lish; 1.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 3.  
 DR PROSITE: PS00082; WD\_REPEATS\_2; 6.  
 DR PROSITE: PS00294; WD\_REPEATS\_REGION; 1.  
 KW Cell cycle; Nuclear protein; Repeat; Ub1 conjugation pathway; Vision;  
 KW WD repeat.  
 FT DOMAIN 4 36 Lish.  
 FT DOMAIN 41 84 F-box-like.  
 FT REPEAT 353 392 WD 1.  
 FT REPEAT 408 447 WD 2.

```
FT REPEAT 449 488 WD 3.
FT REPEAT 491 531 WD 4.
FT REPEAT 532 571 WD 5.
FT REPEAT 574 622 WD 6.
FT REPEAT 625 664 WD 7.
FT REPEAT 666 700 WD 8.
FT DOMAIN 206 323 Ser-rich.
FT MUTAGEN 16 16 L-Q; induces ectopic S phases
within the peripheral and central nervous
system.
FT MUTAGEN 510 510 C-Y; In ebi4; induces defects in R7 cell
development in the eye.
FT MUTAGEN 602 602 S-L; In ebiC3; induces ectopic S phases
within the peripheral and central nervous
system.
SQ SEQUENCE 700 AA; 72387 MW; 286D8D07BB79FB7 CRC64;

Query Match 75.5%; Score 2753.9; DB 1; Length 700;
Best Local Similarity 52.8%; Pred. No. 6.6e-47;
Matches 389; Conservative 51; Mismatches 36; Indels 261; Gaps 75;

QY 1 MSISDEVNPLVRYIQESGFHSAPFTFGIKSHISQSNINGALVPPAALISIIQKQYV 60
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MSFSDEVNPLVRYIQESGFHSAPFTFGIKSHISQSNINGALVPPAALITLQKGLYT 60
QY 61 EAEVSINEDGTLFD-GRPIESLSLIDAVMPDV-----VOTRQ-----AYRD---KLA-- 104
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
61 EVEMSVGEDG---EVARPIEGLSLIDAVMPDVKPLKIVKT-EPGKPGAV-DSAP-AGG 114
QY 105 -Q-----Q-----Q-A-----Q-A-----AA--A---A 112
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
115 NONNNAKPEIKBPGTGVAGSAGCNKIAGTTGTSTPTDQSASEVDSGNAANAGTVA 174
QY 113 ----A-----A---AA--ASO--Q-----GSAK--N-G-EN---T-- 130
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
175 GNNAGCAGNOASTGGSNSTPTAGDLAAPGASOKKQNSNEAGSSSGNAGNANATSTD 234
QY 131 -A-----NG-----EE-----N-----G-AH---T- 141
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
235 AASSTSTNGNSSSSVEQPTSGLTAGGTVTSNPDAAAGSASTATGSKAPSGAVTIR 294
QY 142 --A--NN-----H-----T-DMEV--DGD--VEIPPNKAVVLRG 167
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
295 VQAQGNVQSGSSNAQSSAPSGTISSTSGGAGTAAAL-VPMDIDENIBIPESKARVLG 353
QY 168 HESEVFICAWNPVS-DLLASGSDSTARINLSE-NSTSGTOLVLRHCIREGQDVPSN 225
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
354 HESEVFICAWNP-SRDLASGSGDSTARIDWMSDAN-T-NSNQLVLRHCIOKGAEPVSN 410
QY 226 KDVTSLDWNSEGTLLATGSDGPARIW-TKDGNLASTLGOHKGPFPALKWNKK-GNFILS 283
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
411 KDVTSLDWNCDGSLLATGSDGVARIWKT-DGRLASTLGOHKGPFPALKWN-KCNYILS 468
QY 284 AGVDKTTIIDAHTGE-AKQOPFPHPAPALVDVQSNNTFASCTD-MCHIVCKLG--QD 339
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
469 AGVDKTTIIDAHTGCT-QQAFHAPALVDVQSNNTFASCTDQD-IVHVCRLGNE- 525
QY 340 RPLKTFQHTNEVNAIKWPTGNLLASCDDMTLKIMSKQD-NCVHDLQHNKEIYTIK 398
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
526 -PIKTPKHTNEVNAIKWCPQGOGLLASCDDMTLKIMSMNRDCC-HDLQHSKEIYTIK 583
QY 399 WSPGTGTTNNPNANLMLASGSDSTVRLWDVDRGICHTLTKHOEVPVSVAFSPDGRYLA 458
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
584 WSPGTGTTNNPNANLMLASGSDSTVRLWDVDRGICHTLTKHOEVPVSVAFSPDGRYLA 458
QY 459 SSGFDCVHIWNTQ-----VCLH-YLNGQVLLNLGRSI---CLY-----TLPHLV----- 500
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
644 SSGFDCVHIWNTQGLV--HSY-KG-----TG-GIFEVC-WNSKGT-K---VGASASD 689
QY 501 --VPLVALIELLV-L-K 514
Db |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
690 GSV--EV-L-D---LRK 699
```

## RESULT 10

```
AAP49011
ID AAP49011 PRELIMINARY; PRT; 7073 AA.
AC AAP49011;
DT 02-MAR-2004 (TReMBLrel. 27, Created)
DT 02-MAR-2004 (TReMBLrel. 27, Last sequence update)
DT 02-MAR-2004 (TReMBLrel. 27, Last annotation update)
DE Orf1ab polyprotein.
OS SARS coronavirus ZJ01.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Nidovirales;
OC Coronaviridae; Coronavirus; Human coronavirus (strain SARS).
OX NCBI_TaxID=230471;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZJ01b;
RX MEDLINE=22889812; PubMed=14527350;
RA Li L., Wang Z., Lu Y., Bao Q., Chen S., Wu N., Cheng S., Weng J.,
RA Zhang Y., Yan J., Mei L., Wang X., Zhu H., Yu Y., Zhang M., Li M.,
RA Yao J., Lu Q., Yao P., Bo X., Wo J., Wang S., Hu S.;
RT "Severe acute respiratory syndrome-associated coronavirus genotype and
its characterization";
RL Chin. Med. J. 116:1288-1292(2003).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ZJ01b;
RX PubMed=14733771;
RA Wang Z.G., Li L.J., Luo Y., Zhang J.Y., Wang M.Y., Cheng S.Y.,
RA Zhang J.J., Wang X.M., Lu Y.F., Wu N.P., Mei L.L., Wang Z.X.;
RT "Molecular biological analysis of genotyping and phylogeny of severe
RT acute respiratory syndrome associated coronavirus.";
RL Chin. Med. J. 117:42-48(2004).
DR EMBL; AY286320; AAP49011.4; -.
KW Polypeptide.
SQ SEQUENCE 7073 AA; 790251 MW; 96595991DACA6EA9 CRC64;
```

Query Match 71.2%; Score 2599.6; DB 2; Length 7073;  
Best Local Similarity 8.3%; Pred. No. 1.4e-41;  
Matches 464; Conservative 38; Mismatches 5; Indels 5074; Gaps 381;

```
QY 1 M-----S-----S-----DEV-----N----- 9
Db 297 MGRIRSVYVPASPOECNNMHLSTLMKNCDEVSMQTCDFLKATCCHCGTENLVIEGPTT 356
QY 10 ----- 12
Db 357 CGVLPNVAVKPCACQDPETGPEHSVADYNHNSIETRLRKGGRTRCFPGCVFAYGVC 416
QY 13 Y--R-Y-----L-Q--E--S-----G-FSH----- 23
Db 417 YNKRAVVPASADIGSGHTGITGDNVETLNEDLLILSRERVNIIVGDF-HLNEVAI 475
QY 24 -----SAF-----T----- 27
Db 476 ILASFASASTAFIDTILSKLDYKSFKTIVSCGNYKVTGKPKVKAANNIGQORSVLPLCG 535
QY 28 F-----G-IKS-----H-----IS-OS-----N-- 38
Db 536 FFSQAAGVIRISIPARTLDAAHNSIPDLQRAAVTILDGISEQLRLVDAMVYTSDDLTVNSV 595
QY 39 -I-----N--GALVP--P-----A-----A-----LIS- 51
Db 596 IIMAVYTGGLVQQTSQLNSLLGTTVEKLRPIPEWIEAKLSAGVEFLKDAWEILKFLITG 655
QY 52 ---I---IQ-----K-----G-----L-----Q--- 58
Db 656 VFDIVKGQIQVASDNIKDCVKCFIDVVKALEMCIQVITAGAKRLSLNLGEVFIQSKG 715
QY 59 -Y-----V-----V-----EA-E--V-S--- 65
Db 716 LYRQCIRGKEQLQLLMPLKAPKEVTFLEGSDHDTVLTSEVVLKNGEALPTFVDSFTN 775
QY 66 -----I-----N-----ED----- 69
```

Db 776 GAIVGTPVCVGLMLLEIKDEQVICALSPGLLATNNVFRLLKGGAPIKGVTFEGDTWVEVQ 835  
QY 70 -----GT-----  
Db 836 GYKNVRITFELDERVDKVLNEKCSVYTVESGTEVFAPCVAVBAVVKTLQPVSDLLTNMG 895  
QY 72 -----LFD-G-----R-----P-----I-----  
Db 896 IDLEMSVATFYLFDDAGEENFSRMVCSFYPPDEBEEDDARCEBEIDETCSHEYGTED 955  
QY 79 -----E-S-----L-----S-----M-----  
Db 956 DYQGLPLEFGASAEIVRVEBEEDWLDDETEQSEIEPEPEPTPEBPVNOFTGYLKLTDN 1015  
QY 84 -----I-----D-----A-----V-----  
Db 1016 VALKCDIVKEAQSANPMWIVNAANTHLKHGGVAGALNKATNGAMQKESDDYIKLNGPL 1075  
QY 89 -----P-----D-----  
Db 1076 TVGGSCLLSGHNLAACKLHVGNLNAgedIQLLKAAYENFNSODILLAPLLSAGIFGAK 1135  
QY 91 -----V-VOT-R-Q-----QA-Y-----R-----  
Db 1136 PLQSLQVCVQTVRTQVYIAVNDKALYEQVMDYLDNLKPRVEAPKQEEPPNTEDSKTEEK 1195  
QY 101 -----D-----D-----KL-----A-----Q-----  
Db 1196 SVQKPVDRPKIKACIDEVTTTLEETKFLTNKLLFADINGKLYHDSQNMRLGEDMSFL 1255  
QY 106 -----Q-----  
Db 1256 EKDAFVWGVDIITSGDITCVIIRSKAGGTMLSRALKKVPVDEVITTPGGCGAGYTL 1315  
QY 109 -----A-A-----A-A-----A-A-----  
Db 1316 EEAKTALKKCSAFYVLPSEAPNAKEBELLGTVSNLREMLAHAEETKLMPICMQDVRAIM 1375  
QY 118 AS-Q-----Q-----G-----S-----A-----K-----N-----G-----  
Db 1376 ATTQRYKIGIKIQEGIVDYGRFFYTSKEPVASITTKLSNLEPLVIMPIGYVTHGFNL 1435  
QY 128 -----E-----N-----N-----T-----A-----  
Db 1436 EEAARCMRSLKAPAVSVSSPDATVYNGYLTSSSTSEHFVETVSLAGSYRDSYSGQ 1495  
QY 132 -----N-----N-----  
Db 1496 RTELGVFEFLRGDKIVVHTLESFVEPHLDGEVLSDKLKLSLREVKTIKVFTVDNTN 1555  
QY 137 -----CA-----  
Db 1556 LHTQLVDMSTMVYQQGFPTYLDGADVTKIKPHVNHGKTFPVLPSDDLRLSEAFYHYTL 1615  
QY 139 -----HT-----  
Db 1616 DESFLGRYMSALNHTKKWFPQVGGLTSLKWADNVCYLSVLLAQOLEVKNAPALQEA 1675  
QY 141 -----I-----AN-----N-----H-----  
Db 1676 YRARAGDAANFALILAYSNKTVGBELGDVRETWTHLLQHANLESASAKVLNVVCKHCGOK 1735  
QY 146 T-----D-----MM-----  
Db 1736 TTTLTGVEAVWVMTLSYDNLTGVSIPCVGRDATQYLVQBSSEFVMSAPPABYKLOQ 1795  
QY 150 -----E-----VDG-----DV-----  
Db 1796 GTFLCANEYTGNCQGHYTHITAKETLYRIDGAHLTMSEYKGPVTDVFKYTSYTTIK 1855  
QY 156 -----BT-----P-PN-----  
| |

Db 1856 PVSUKLDGVITYTEIEPKLDGYKKNAYYTEQPIDLVPTQPLPNASFDNFKLTCSNTKFA 1915  
QY 161 -----K-A-----V-----  
Db 1916 DDLNQMTGTFKASRELSVTFPDLNGDVVAIDYRHYSASFKKAGAKLLHKPIVHMHQAT 1975  
QY 165 -----LR-----  
Db 1976 TKTTFPNTWCLRLCLWSTKPDVTSNFEVLAVEDTOGMNLCESQOPTSEEVENPTIQ 2035  
QY 167 -----GHE-----S-----E-----  
Db 2036 KEVIECDVKTTEVGVNVLKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNELSLALG 2095  
QY 172 -----VF-----  
Db 2096 LKTIATHGIAAINSVPSWKILAYVKPFLGQAAITTSNCAKRLAORVFNMYVPTLLFQ 2155  
QY 174 -----I-----C-----A-W-----  
Db 2156 LCTFTKSTNSRIRASLPITIAKNSVKSAKLCLDAGINVYKSPKSLFTIAMWLLLSI 2215  
QY 178 -----N-----  
Db 2216 CLGSLICVTAAGVLLSNFGAPSYCNGVRELYLNSSNVTTMDPCEGSFPSCICLSGLDSL 2275  
QY 179 -----P-----V-----S-----DI-----LA-----S-----G-----S-----  
Db 2276 DSYPALETIQVTISSYKLDLTILGLAAEWLAYMLFKFFYLLGLSAIMQVFFGYFASHF 2335  
QY 190 -----D-----ST-----  
Db 2336 INSNLWMLFIISIVQMAPVSAMVRVIFPASFYIHWKSYVHMDGCTSTCMCYKXNRA 2395  
QY 194 -----R-----I-----WN-LEEN-ST-SGST-----  
Db 2396 TRVECTTIWNGMKRSPVYVANGRGFCFTHNNCL--NCDTFTGTFISDEVARDLSLQ 2453  
QY 208 -----Q-----L-----V-----L-----BH-----  
Db 2454 FKRPINPTDQSSYIVDSVAVKNGALHLYFDKAGQKTYERHPLSHFVNLOLNRANNTKGS 2513  
QY 214 -----C-----I-----  
Db 2514 PINVIVFDGSKCDBSASASVYYSQLMCQPIILLDQALVSDVSDGSTEVSVMQFDAYVD 2573  
QY 216 -----R-----  
Db 2574 TFSATFSPWMEKALKALVATAHSELAKGVALDGLSTFVSAARQGVVDTDVTDKDVIECLK 2633  
QY 217 -----E-----G-----G-----Q-----  
Db 2634 LSHSDLEVTGDSNNFMTYNKVENMTPRDLGACIDCNARHINAQVAKSHNVSLIWNVK 2693  
QY 221 -----  
Db 2694 DYNLSEQLKQIRSAAKNNIPFRITCATTRQVNVNITTKISLKGKIVSTCFKLMLKA 2753  
QY 222 -----V-----P-----S-----N-----K-----D-----VT-----S-----LD-----  
Db 2754 TLLCVLAALVCYIVMPVHTLSIHDGYTNEIIGYKATODGVTRDIISTDDCFANKHAGFDA 2813  
QY 233 W-----N-----S-----E-----GT-----  
Db 2814 WFSQRGSYKDKSCPVAAIITREIGFVPGVLRAINNGDFLHFLPRVPSAVGNIC 2873  
QY 239 -----I-----LA-----T-----G-----SY-----  
Db 2874 YTPSKLIEYDFATSACVLAECTIPFKDAMGKPVPCYDNLNLEGSIYSSELPRDTRYVL 2933  
QY 246 -----DG-----F-A-----R-----I-----W-----  
Db 2934 MDGSIIFPNTYLEGSVRVVTTFDAFYCRHGTCERSEVGICLSTSGRWVNLNNEHYRALSG 2993  
| |

QY 253 ---TK---D--- 255  
Db 2994 VFCGVDMNLIANI FPLPQVGDALDVSASVWAGGIIAIIIVTCAAYYPMKFRVFGCEYNH 3053  
QY 256 ---G---N---LA---S--- 260  
Db 3054 VVAANALLFLMSFTILCLVPAYSFLGVSVFYLTYFTFTNDVSLAHLQWPAFSPIV 3113  
QY 261 ---T---L---G--- 263  
Db 3114 PFWITAIYVFCISLKHCHWFNNYLRKRVMFNGVTFSTPEEAALCTFLNKEMYLKLRSE 3173  
QY 264 ---Q---H---K---G---P--- 268  
Db 3174 TLLPLTOYNNRYLALYNKYKYFSGALDTSYREAAACHLAKALNDFNSGADVLYQPPQTS 3233  
QY 269 I---F---AL---K---W--- 274  
Db 3234 ITSAVLQSGFRKMAFPKGVGCMVQVTCGTTTLNGLMLDDTVYCPRHVICTAEDMLNPN 3293  
QY 275 ---N---K--- 276  
Db 3294 YEDLLIRKNSHSLVQAGNVQLRVIGHSMQNCILLRLKVDTSNPKTKYKRVRIQPGQTS 3353  
QY 277 ---KG---N---F---I---L---S---A 284  
Db 3354 VLACYNGSPGVYQCAMRPNHTIKGSLNGSCSVGFNDYDCVSCYMHMELPTGVHA 3413  
QY 285 G---VDK---TTI---I---W---D--- 294  
Db 3414 GTDLEKFGYFVDRTQAAGTDTITLNVLAWYAIVNGDRWFLNRTTLNDNFNV 3473  
QY 295 A---H---TG---E--- 299  
Db 3474 AMKYNEPLTQDHVIDILGPSAQTGIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTP 3533  
QY 300 ---AK---Q---O---F---PF--- 306  
Db 3534 FDVVRQSGVTFQGEKKIVKGTTHWMLLTFTLSLLILVQSTOWSLPFFVYENAFIPFTL 3593  
QY 307 ---H---S---A---PA--- 311  
Db 3594 GIMAIACAMLLVKHGHAFCLFLPLSLATVAYFNVMPVMPASVMRMIMTWLEADTSLG 3653  
QY 312 ---L---D---V---D---W---Q--- 317  
Db 3654 YRLKDCMYASALVLLIIMTARTVYDDAARRVWTLNMTLVYKYVYGNALDQAIMWAL 3713  
QY 318 ---SN---NT--- 321  
Db 3714 VISVTSYSGVTTIMFLARAI VFCVVEYYPPLLFTIGNTLQCLIMVYCFLYGCCCYFGL 3773  
QY 322 FASC---ST---D--- 328  
Db 3774 F---CLLNRYFRLTGVVDVLSVTOEPRYMSQGLPPKSSIDAFKLNILKLGIGGKPCIK 3831  
QY 329 ---M---CI---H--- 332  
Db 3832 VATVQSKMSDVKCTSVLLSVLQQLRVESSSKLWACQVQLHNDILLAKDTTEAFERQVSL 3891  
QY 333 ---V---CK---L---G--- 337  
Db 3892 LSVLLSMQGANVINRICEMLNDRATLOIAESFSLPSYAAYATAQEAQEAVQAVANGDSE 3951  
QY 338 ---O--- 338  
Db 3952 VVLKLLKSLNVAKSEFDDAAMQKLEKMAQAMTQMYQARSEDKRAKVTSMQTMFLP 4011  
QY 339 ---D---R---P---I---K---TF--- 345  
Db 4012 TMLRKLNDLANNIINNARDGCVPLNIIPLTTAAKLMVVVVPDYGTYKNTCDGNTFTYASA 4071

QY 346 ---Q---GH---T---NE---V--- 352  
Db 4072 LWEIFQVVDADSKHVQLSBIINMDSNLAWPLIVTALRANSVAVKLQNNELSPVALROMSC 4131  
QY 353 ---NA---I---KW---D---P--- 359  
Db 4132 AAGTTQTACTDDNALAYNNNSKGRFVLLALSDHQLKWARFPKSGDTGTIYTELBPPCR 4191  
QY 360 ---T---G---NL---LA---S---CS--- 368  
Db 4192 FVTDTPKGPKVLYPIKGLNNLRGMVLGSLAATVRLQAGNATEVPAHSTVLSFCAPAV 4251  
QY 369 D---D---M---T--- 372  
Db 4252 DPAKAYKDLASGGQPIITNCVRLMCTHTGTGOAITVTPEANMDQESFGGASCCLYCRCHI 4311  
QY 373 ---LK---I---W---S---M--- 378  
Db 4312 DHPNPKGCDLKGKVVQIPTTCANDPVGFTLRNTVCTCGMMKGYGSCDQLREPLMQSA 4371  
QY 379 ---K---Q--- 380  
Db 4372 DASTFLNRVCGVSAARLTPCGTGTSTDVVYRAFDIYNEKAVGPAFLKTNCCRFQEKDEE 4431  
QY 381 ---D---N---C---VHDL---Q--- 388  
Db 4432 GNLLDSYFVVKRHTMSNYQHEETIYNLVKDCPAVAVHDFPKPRVDGDMVPHISRQLTKY 4491  
QY 389 ---OH---N---KEI--- 394  
Db 4492 TMADLVYALRHDFEGNCDTLKEILVTYNCCDDYFNKKDWDYFVENPDILRVYANLGERV 4551  
QY 395 ---Y--- 395  
Db 4552 ROSLLKTVQFCAMRDAGIVGVLTLDNQLNGWNVDFGFEQVAPCGGVPIVDYSYSLLM 4611  
QY 396 ---T---IKW--- 399  
Db 4612 PILTLTRALAASHMDADLAKPLIKWDLKLYDFTBEERLCFLDRYFKYWDQTYHPNCINCL 4671  
QY 400 ---S---PT---GP--- 404  
Db 4672 DDRCILHCANFNVLSTVPPTSPGFLVRKI FVDGVFPFVSTGYHFRELGVHVNQDNLH 4731  
QY 405 ---G---TNN---P--- 409  
Db 4732 SSRLSFKELLVYAADPMAHAASGNLLDKRTCTCFVAALTNVAVFOTVRPGNFKDFYDF 4791  
QY 410 ---NA---NL---M---L--- 415  
Db 4792 AVSKGFFKEGSSVELKHPFFAQDGNAAISDYDYRYNLTMTCDIRQLLFVVEVVDKYFDC 4851  
QY 416 ---A---SASF---DS--- 422  
Db 4852 YDGGCINANQVNNLNDKSAAGFPFNKWKARLYYDSMSYEDQDALPAYTKRNVIPITLM 4911  
QY 423 ---TV---R---L---W--- 427  
Db 4912 NLKYAISAKNRARTVAGVISCSTMNTRQHQKLLKSIATRGATVTVIGTSKFGYGGWHNML 4971  
QY 428 ---DV---DR--- 431  
Db 4972 KTVYSDVETPHLMGWDPYKCDRAMPNMLRIMASLVLRKHNTCCNLSHRFPYRLANECAQV 5031  
QY 432 ---G---IC---I--- 435  
Db 5032 LSEMVMCGGLYVPGGTSSGDATTAYANSVENICOAVTANVALLSTDGNKIADKYVRN 5091  
QY 436 ---H---TLT---KH--- 441  
Db 5092 LQHRLYECLYRNRDVEHFYFAYLRKHFSMILSDDAVVCYNSNYAAQGLVASIKNF 5151  
QY 442 ---Q---E---P---VY---S---V---A--- 449

```
Db 5152 KAVLYYQNNVFMSEAKWTETDLTKGPHFCQSQHTMLVKQDDYVLPYPDPRIILGAGC 5211
QY 450 -----P-S-----P-----D-----G-----454
Db 5212 FVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNOEYADVPHLYLQYIRKLHDELTHMLDM 5271
QY 455 -----RY-----L-----A-----S-----GS-----P-----462
Db 5272 YSVMLTNDTSRIWEPEFYEMATPHTVLQAVGACVLCNSQTSLRGACIRRFPLCKKC 5331
QY 463 -D-----K-----C-----Y-----H-----467
Db 5332 YDHVISTSHKLVLSVNPVYCNAPGCDVTDVTLVLGMSYYCKHPPIISFPLCANGQVF 5391
QY 468 -----RY-----I-----W-----NTQVC-----474
Db 5392 GLYKNTCVGSDNVDFNAIATCDWTNAGDYILANT--CTERLKLFAAETLKATEFTFKLS 5449
QY 475 -----LH-----476
Db 5450 YGIATVREVLSDELHLHLSVEVGKPRPLNRNYVFTGYRVTKNSKVQIGEVTFEKGYGDA 5509
QY 477 -----Y-LN-G-----Q-V-----L-----LN-----485
Db 5510 VVYRGTTTYKLVNGDYFVLTSHTVMPLSAPTLVPQEHYVRIITGLYPTLNLISDEFSSNVAN 5569
QY 486 -----L-----G-----R-----488
Db 5570 YQKVGQKYSTLQGGPPTGKSHFAIGLALYPSARIIVYTACSHAADVADALCEKALKYLPID 5629
QY 489 --S--I-----C-----L-Y--T--LP-----496
Db 5630 KCSRIIPARAVECFDKFKNVSTLEQVFCVTNALPETTADIIVFDEISWATNYDLSVN 5689
QY 497 -----H-----497
Db 5690 ARLRAKHVYVIGDPAQLPAPRTLLTKGTLPEYFNSVCLMKTIGPDMFLGTCRCCPAEI 5749
QY 498 -----H-----L-----V-----501
Db 5750 VDTVSALVDYDNKLKAHKDKSAQCFKMFYKGIVITHDVSSAINRPQIGVVRFLTRNPAWRK 5809
QY 502 --I-PL-----VA-----L-----I-----EL-LVL-----513
Db 5810 AVFISPNSQNAVASKILGPTQTVDSQSGSEYDVIPTOTTETAHSCNVNRFNVAITRA 5869
QY 514 K 514
Db 5870 K 5870

RESULT 11
Q8NJX1 PRELIMINARY; PRT: 20925 AA.
ID Q8NJX1
AC Q8NJX1
DT 01-OCT-2002 (T-EMBLrel. 22, Created)
DT 01-OCT-2002 (T-EMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)
DE Nonribosomal peptide synthetase.
GN Name=pepI;
OS Trichoderma virens.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=29875;
RN [1]
RP MEDLINE=22037864; PubMed=11909873;
RX Wiest A., Grzegorski D., Xu B.-W., Goulard C., Rebuffat S.,
RA Ebole D.J., Bodo B., Kenerley C.;
RT "Identification of peptaibols from Trichoderma virens and cloning of a
RT peptaibol synthetase.";
RL J. Biol. Chem. 277:20862-20868(2002).
```

```
[2]
RN SEQUENCE FROM N.A.
RP Wiest A.S., Grzegorski D., Xu B.-W., Zhou W., Ebole D.J.,
RA Kenerley C.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
DR EMBL: AF469045; AAM78457.1; -.
DR HSSP: O30409; 1DNY.
DR GO: GO:0048037; F:cofactor binding; IEA.
DR GO: GO:0016740; F:transferase activity; IEA.
DR GO: GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR010071; AA adenyl_dom.
DR InterPro: IPR009081; ACP_like.
DR InterPro: IPR001227; AC_transferase.
DR InterPro: IPR000873; AMP_bind.
DR InterPro: IPR001242; Condensatn.
DR InterPro: IPR000794; Ketoacyl_synth.
DR InterPro: IPR006162; Ppantne_S.
DR InterPro: IPR006163; Pp_bind.
DR InterPro: IPR010080; Thioester_redct.
DR Pfam: PF00698; Acyl_transf_1; 1.
DR Pfam: PF00501; AMP-Binding; 18.
DR Pfam: PF00668; Condensation; 27.
DR Pfam: PF00109; ketoacyl-synt; 1.
DR Pfam: PF00550; PP-binding; 19.
DR PRINTS; PRO0154; AMPBINDING.
DR TIGRFAMs; TIGR01733; AA-adenyl-dom; 17.
DR TIGRFAMs; TIGR01746; Thioester_redct; 1.
DR PROSITE; PS00075; ACP DOMAIN; 19.
DR PROSITE; PS00455; AMP_BINDING; 18.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN 11.
DR PROSITE; PS00012; PHOSPHOPANTHETHEINE; UNKNOWN 11.
SQ SEQUENCE 20925 AA; 2320297 MW; C6603A269D56FE11 CRC64;

Query Match 71.2%; Score 2598.3; DB 2; Length 20925;
Best Local Similarity 8.2%; Pred. No. 1.2e-40;
Matches 465; Conservative 41; Mismatches 6; Indels 5177; Gaps 385;

QY 1 M-----S-SS-----D-----E-----V-----8
Db 8843 MISYEQYGLQNTSKISSHAMEACDFGLIIQIEQLAYRGDDALLVAEASHITNETSAL 8902
QY 9 -N-F-----LVY-----R-----Y--LQE-----18
Db 8903 QNVFSYPLVIOCHLYDNTINLMYDSQALSSRMEALSHHFDNAVQELASKDNLLGSI 8962
QY 19 S-G--F-----S-----H-----S-A-----FT-----27
Db 8963 SLAGRWDFFQALKLNSDSPDAVERCFHEMLDEIALVRGDSLAIISGWDKSFYREMAETT 9022
QY 28 -----F-----G-----I-KSH-----I--34
Db 9023 RIAHLVTDYGVKVGDIHVCPEKSAWFTIATLAINKAGAAWSTLDPSPHPTERYQKINQ 9082
QY 35 --SQ-----SN-----46
Db 9083 TGSQIALSSPVNSNRCSALLPNVIELTSELDSKLAQNVWSTSRPAVNTPRDAAYVLFT 9142
QY 39 -----I-----N-----GA-42
Db 9143 SGSTGVKGVWIEHGSLSCTSIQSLSKALDFNEEFVLFQSSYSFDIILFEIGSTFLTGC 9202
QY 43 -----L-----VP-----P-----46
Db 9203 LFPVSWDEQNMELVEYIRKHQLTFFMLTPTLARTIRPEDVPSVDMVLVAGEAPTRDILDI 9262
QY 47 -----A-----A-L-----I--S-----51
Db 9263 WFGKRLANGWPTGTECSVIACLHQWTSVDESVPKVGRIPIGGSCWIVDPEDATCWAPLGT 9322
QY 52 --I-IQ-----K-----GL-Q-----Y-----V-EA--E 63
```

```
Db 9323 GEIVIOQRNLLREYLSDPVKATATITGLPQWPKRDSIHWDRFYLTGDLGFINEAGNLE 9382
QY 64 -----VSI-----N-E-----D-G-TL----- 72
Db 9383 YCTRKTQVKIRQORLEGEIEHHIQANLESVRQVAVDVVKS DAGSTLVAFVSFSDATQ 9442
QY 73 -----F---DG-----R---P---I----- 78
Db 9443 TVTDTRVIANGPSTLDGDFQATISRLIGILSTLMPRYMVPASAFIPCAVKGEBFQHTGGRY 9502
QY 79 -----E---SLS-----L---ID----- 85
Db 9503 KKLRELAASLOEDMAAYSLANBEKVAPOPAMESRIQKIWAQVLNIDSIGRDDSFLQI 9562
QY 86 -----AV-----M---P---D---90
Db 9563 GGDVSUVRILSVARDANIRLTVDIFDDPRLLAVAAKATEDWRDGLMTQTEPFVLLDG 9622
QY 91 -----VV---QT-R-----Q-----A---YR-----D 101
Db 9623 SVKNIVLSQNIROYSLPADKELEDAYPCTKLQEGLMALAVKQPGSVIAKFLYRISPNVD 9682
QY 102 -----K-----L-----AQ-----Q-----QA 108
Db 9683 VSHLRASWEETVRLPLNLRTRIFTVNMSVQAVFKDDVSNQSTGTSLSRYMCGSTENFQM 9742
QY 109 A-----A---A-----A-----AA----- 114
Db 9743 AYGSPLARVALIEDGGHTVFILSMHHA VFDGWMGRVFMVAFHISIFNHRTHTVVEPVRFV 9802
QY 115 -----A---AAA---SQ-----Q----- 121
Db 9803 KYTLAIDSDAAADYMSQEGARQSIFPPTLSSVHEKKENSTLTLEKVIDLPSTKSSIT 9862
QY 122 -----G---S-----A-K----- 125
Db 9863 TATMLRAAWAII LARYCSDSDITFGATISGRQAPVPLADMTGPAITVPVRVVRINRNL 9922
QY 126 -----N-----G----- 129
Db 9923 ISDYLOAIQRANDMTIPFQFGLQNTSKLSDDAKADCFSSLLVIOIQSLSVVDENADA 9982
QY 130 -----T-AN-----G---E-----EN----- 136
Db 9983 IFVQAEVEKEIGTVQNYFSYPLAVIQGHVENFINLVLYDSNVLSNQITALTSHQFESV 10042
QY 137 -----G---A-----H-----T-----I--- 141
Db 10043 MKQLASQSDMKLGSVMSASDWDLEHSMRQNRDIPDIVDSCIHELIOYQAVTQPDAPAIVS 10102
QY 142 -----A-----A-----N-----N----- 144
Db 10103 WDRDFTYKQLEASNLRLHLVNKNVKNPKDDLIPVFFERNAMVFAITAINKAGAAWVPL 10162
QY 145 -----H----- 146
Db 10163 DPSPHVLRLRQLISQGTTLALSSSANAVLCSFLVRKVVEVNAELDNKLLATESAHGPV 10222
QY 147 D-----M----- 148
Db 10223 VDVSSRNAAYVLFTSGTGIPKGLIMEHGSVCTSQVAIAKRLGLNSKVRILQFAAFVFDL 10282
QY 149 ----- 148
Db 10283 SIGEIVGLISGACICVPSEHIRKNSIANFINRQGITWTYLTPTSFVRTIKASEVNPVKLL 10342
QY 149 -----M-----E-----V 151
Db 10343 LLAGAEVPRDIFATWFGKRLRLINGWPAETCCFSTLHEWAI CGRESFYCCRPVGGFCWIV 10402
QY 152 D-----G-----DV---E---I---P---P-----NK 161
Db 10403 DPENPHRLAPTALGEVITQGPILREYLSVDVDRTEAAVIKSLPDWAPFREQSSGRFNK 10462
```

---

```
QY 162 -----AV-V-L----- 165
Db 10463 SGDLGVYNPDGTIEFSSRKDTQVKIRGLRVELGEVEHVAQVALDGVHQIATAVDVFKGDNGT 10522
QY 166 -----RG-----HE---S-----E-----V-----FI-C-- 175
Db 10523 QPRGYFSDDESQIHEAHPSGFPQAI DEKLOARLTVAIGKLNIALPRYMIPTLFPCKY 10582
QY 176 -----A----- 177
Db 10583 MFSITSTKLDRELRRRTLLTQSELAMFSYRCNKTPTPTMESLQRIWSEILHPLDS 10642
QY 178 -----NP---V-----S 181
Db 10643 IGRDSDFLGLGGDSITAIHLGSI SRDQGISITAKDVPDDPRLWDIASKARELDEAQQQS 10702
QY 182 D-----L-----LAGS-----G----- 189
Db 10703 PDISPFSFLIGETRELAAGSVKQKLLTSQTIEDAYPCTSKI PGSIYIAKYVYRLPHQV 10762
QY 190 D-----S---T-A-----RI-----W-NL-S---E-N-----S 202
Db 10763 DIGRFKSWEHVALHPILRTRIIMIKOTCVQLVVKDSVTWENLDSGDLDEAVNATHSYS 10822
QY 203 -T-----SG-----S- 206
Db 10823 MTYGSPLSRYSIHESKSGDKYFLWTAHHSIHGWSVPIFNTLYQAYKGLGKPKAYS 10882
QY 207 -T-----OL----- 209
Db 10883 FIKYTWMEHQKAAGDYWRQLOQNAKASFPKGTGLAAKSIWTKRTRVMTSTLKFSSASSNE 10942
QY 210 -----V---L-----R----- 212
Db 10943 IATKASVIRAAWAVILARYCESDDVCFGATISGRQASVFCLEMGAPAVATVPVRVQLDN 11002
QY 213 -----H-----C-----I-R----- 216
Db 11003 DQESKFLQNIQSOAHMVPYEQYGLQSI AKLGNDRDARDADFTSLVLIQIQRQESVQS 11062
QY 217 ----- 216
Db 11063 DDALLIPVETELEDILQSYNYPLVLOQHIYNDYADLVLYDSTVTVSEPMQALCHFFNN 11122
QY 217 -----E-GG-----Q---D---V-----P-- 223
Db 11123 IVQQLAAEDGGKLGDISIASSFDLELAORSNGDGPQIIDDCHIHLIERQAKORPNRPAI 11182
QY 224 -----S-----NK----- 226
Db 11183 DAWDGKFTYSELDRPTANRLAHLVHDYAVKVGDIVHVCFEKSQWYFVALAVNKAANA 11242
QY 227 -D-----V-----T-----S-LD--W---NSE- 236
Db 11243 PFDPAHPQORLKAVASOTGAKLALASTANTRLCEQVDCVWEVSSTLDKNLWTTYNIEK 11302
QY 237 G-----T-----L-----A-----T----- 242
Db 11303 GPDINVTMDAAAYILFTSGSTGVPKGIVMQHGALCTNQAAALSGWLGFHTVRMLQFSSFV 11362
QY 243 -----G-----S-----Y----- 245
Db 11363 FDUVSVEIVQALMNGACVCPSEHRLNLSDFVRDPFNVTWAYLTPTSFRTLKPKDFPPL 11422
QY 246 -----D---G---F-----A-----R----- 250
Db 11423 KLLLLAGEPTQDVLDTWFLGNTRFINAWGPAETCVYNTLYEWQSNSTESSPLKLGRAVGA 11482
QY 251 -IW----- 252
Db 11483 YIWWVDVENPORLAPTGCIGLIIVQGPILLKEYLADPEKTAATAATVTEPEWAPRQSTTW 11542
```

QY 253 ---T---KD-----255  
Db 11543 NRFRYTGDLGYDHDGMLHFASRKDTQVKIRGLRVELGEVEHRIQGSLDGVRQVANDVFK 11602  
QY 256 ---G-NL---AS--T---LG-----263  
Db 11603 TERGANLVAYLCTDKTTPDQNTDFGKDVFPASIDFELQINLGKMLTELNSSIPTWIP 11662  
QY 264 -----Q-----264  
Db 11663 TMFIPCAYPFITSSKLDRLVKRLRLTAELSODOLEAYSLLVNKBQAPETEMEIRLOKMA 11722  
QY 265 -----H-----K--G-----267  
Db 11723 EVLDLPEASIGRDNFMFRIGGDSIAAIRLVSMARDAGISLTVNDIFPDARLISVATKAID 11782  
QY 268 -----PI-----FAL-----272  
Db 11783 NDESHLWAPIEPFSLTSGIEDFALPSELSDGOFIEDAYPCSKLOEGLMALAKQPGSY 11842  
QY 273 -----K--W-----N-----275  
Db 11843 IAKYIYKSEHVVDKFKAAWESTVEIASALRTRIIQTGGCSQVWINGDISWDEADGDL 11902  
QY 276 -----KKG-----N--F--I--LS-----AGV 286  
Db 11903 QSLAQSOQLMEGYGTRLCRYTLIRRDNDTYFLWNMHAVIDGLSTQNILGTLFNIYSGV 11962  
QY 287 -----DK-----288  
Db 11963 DVLPPLPNRFIOVMLQLEEDTANYWKNQHNARTIFPPNDVARDKPAATQMLLOSSIE 12022  
QY 289 -----T-TI--I--W-----293  
Db 12023 LPHGLDNSTVIATVIRAAWAIVLARYCDSDDVTFGTTISGROAPIEIMGVPTIATV 12082  
QY 294 -----DA-----295  
Db 12083 PVRVLNRQQLISDFLEGVORQAVEMIAFQYGLQNTAKLGDDARDACDFSSLLVQPIQ 12142  
QY 296 H---T-----GE--A--K--Q-----Q-----F-----304  
Db 12143 HLAGTKDDSLVDASIGEDNAVEAIQNYFSPVLVIQAHLYDDHISFILIYDSSIMLEPQ 12202  
QY 305 -----P-----F-----306  
Db 12203 MVALSVHLOHVMQLTNTPTATLETVSVSSSYDVERALAINAEIPEVIDTCHEMFENRV 12262  
QY 307 H-S--APA-----L-D--V--D-----315  
Db 12263 HLNPLAPAAVADGNYTYAELNSAANKLAHLYKSYSVKLNDLVHVCFDKSAWYIVSIIA 12322  
QY 316 -----W-----Q-----S--N-----N-----320  
Db 12323 INKAGATWPLDPFPHPQRLSIVNQTKSTALASPGNITLCSALVDNVNVEVSALDSML 12382  
QY 321 -----T-----PAS-----C-----325  
Db 12383 PAQGLSPKVSVSSRTAAVILFTSGTGTPKGVVIQHRSLCTNMNAIAKRVRYHLDVRI 12442  
QY 326 -----ST--D--M-----C-----330  
Db 12443 LQFSAYVDFDSILEIINSLQACVCPVSEHIRMNGIVDFIRDMNINWLTPSLRTIN 12502  
QY 331 -----IH-----332  
Db 12503 PIDVPNVELVAGGEAIPRDVFETWVRGRVFRINGWGTETTVVGSIHESFVDESPTIG 12562  
QY 333 --V--C-----K--LG-----Q-----D-----RP 341  
Db 12563 HPVGGFCWIVDPNPNQLLAPTGTILGEIVIOGPTLLHLYDNPKTOBAILYDLPEWAPRP 12622  
QY 342 -----IKT-----F-----Q-- 346

Db 12623 DENNWGRPYKTDGLGFYNANGKIEFSRKDTQVKIRGLRVELGEIEYQVQASVEERQIA 12682  
QY 347 ---G-----H-----TN--E-----V-----352  
Db 12683 VDVIKTDGNSNLVAYLCFNDIEIRQLHNADVNGPSPDLTNLQETLAGAIGKLSVTLPRYM 12742  
QY 353 -----N-----A-----I 355  
Db 12743 IPTFYPCSYMPSITSGKLDRLKELKQTAALSOSELNKFSLHGVSKRAPETPMELQONI 12802  
QY 356 KW-----DP-----359  
Db 12803 -WSKLLSIPSESIGRDDSFLGLGDSIMAIHLVTACREAGVSLTVKEIFDDPRLSAVASH 12861  
QY 360 -----T-----360  
Db 12862 ARSMDAIDQDLVLPFLLSRLREMLVSDTRFQTSLLPSQVIEDAYPCSKLOEGLMAL 12921  
QY 361 -----G-----N-----LL-----364  
Db 12922 SVKQSGSYVAQYVYKLSGVDLVRFKASWERTIQLCANLRTRIVMLDGTVCQLLVDPAP 12981  
QY 365 --A--S--C-----S-----D-----D-----M-- 371  
Db 12982 WDNAPSTDLOCTLNATRESEMTYGSRLNRYALVRDPFEGNHFIVWSSHAVHDGWTLRIMN 13041  
QY 372 TL-----373  
Db 13042 TLYSLYDQGLPNLLPYSAFIRYTNIDKEASRFWTEQMKNAKRATYPPMPRIESHTGV 13101  
QY 374 -----K--I--W-----S-----M 378  
Db 13102 SRMMNMTISFPSPSVKTVNTKATILRGTWAILLARYCDTDDITFGTTVSGRQAPVGLTEM 13161  
QY 379 K-----Q-----D-----N-----C-----383  
Db 13162 PGVWATVIRVLEASQIVAOFLSNIOQATDMIAFEQGLQNI1K1STDADACEFSS 13221  
QY 384 --V---H-DL-----Q-----388  
Db 13222 LVIQPRSHDLTIKGSESSELVASVDARSQEOLIQNYFTYPLVIQGHVFDSEIELLT 13281  
QY 389 -----Q--H-----N-----KE-----393  
Db 13282 YDSTILSEVQMKALSHQFNVANQLVNESNDPLSSITISGEWDLQAKWNVENPELDT 13341  
QY 394 -----I-----YT-----IK-----398  
Db 13342 CIHSLIEQARIRPDAPICAMDGENYQNLNSAANQLAHILKIGIKADDLVHVCFEKS 13401  
QY 399 -W--S-----PT-----G-----POTN-----N-- 408  
Db 13402 VMFVSIIAINKVGAADVPLDPSPQRRLRVQVVGOTLAKFALSSP-TNAALCNKLVHVI 13460  
QY 409 --P-----N--A-----N-----L-M-----LA 416  
Db 13461 EVSPSLIDLSKFCDFGNFPAINVPSSNAYVLFTSGTGTPKGLVMQHGAVCTSQTAIA 13520  
QY 417 --S-----AS--FD-----S--T-----V--RL-W- 427  
Db 13521 KRLSLTPDVRILOFAAVVFDLSIGEIVAPLIHGACVCPSEETRMNGLKEFIRDARINWA 13580  
QY 428 -----DV-----429  
Db 13581 YLTPSFVTRLRPEDVPSLQLLLAGBAVGDRILDWTFGKVRVLINGWPAETCVFSTLHEW 13640  
QY 430 ---D-----R--G-----I--C-----434  
Db 13641 SSIDESPLTIGRPVGYCWIVEAEDSNKLTPIGCLGEVVLQVFTLLREYLAQPORSKETI 13700  
QY 435 I-----H-----TL-----T--K-----440



Db 13701 ITPLPWPAPKQSDAHLWSRFYKSGDLCPYNPNTLFFYSRKDTQVKINGRLVGLGEVH 13760  
QY 441 H-----Q-----E-----P-----VY-----446  
Db 13761 HIRELLEGVRQAVDVLITSETGTQLVSVYICFNDDSQSPSPKASDIYLPDADIQARIT 13820  
QY 447 -----SVA-----F-----450  
Db 13821 SMVGLSVTLPRYMIPTLPIPKYMPFITSTKLDRTKLSITASLGRDLIHVSLNSKK 13880  
QY 451 -----S-----P-D-GR-----YLAS-----G-S-----461  
Db 13881 RVPETEMTRLOAIWSEILNLPIDISGRDSDFLQIGDGSITAIYLVSKAREAGISLIINQ 13940  
QY 462 -FD-----K-C-V-----HI-----468  
Db 13941 IFDSELLAVASKAVLSTEFQBPQPIVFPFSLNKNQTRALVLAGEVRKFCGLNDDHIETD 14000  
QY 469 -----WN-----470  
Db 14001 AYPCTSLQELMALTMKQSGVYAKYVKLSTFVDMERFQAAWNRMTMELCGNMRTRIVLL 14060  
QY 471 -T-----Q-----V-----C-----L-----475  
Db 14061 NGTPIQLLKEDNQWQSLNDTLASITNSSRDLKMGYGAPLCWYGVLENNNNKYFWWSAH 14120  
QY 476 H-----YL-----N-----G-----480  
Db 14121 HSIYDGVWVRILLTYLSTDTVPTLPYSGFIKYNMELDSMALADFWRELQSGSKRA 14180  
QY 481 -----Q-----VL-----483  
Db 14181 VFPARTSPSSSSQIPKSIISLEQAKQSIITKASILRAAWAIVLARYCYDDTDDVSFGTT 14240  
QY 484 -----LN-----L-----G-----487  
Db 14241 VSGRHAPVAGLETMPGPMIATVPVRVHLNRSATKSQFLTEIQOAYEMVYBPQFGLONIS 14300  
QY 488 -----RSIC-----L-Y-----T-----LP-----496  
Db 14301 KLSRDARTCDFSLLLVIQPPATTISEKADTNILYIGDEBQSLLTDAMHNYFNPLVII 14360  
QY 497 -----H-H-----LV-----500  
Db 14361 LNTFEDHILQRFENSEVLTEARVSALSNDHVVKKLAETDETLESVNLVGDWDIQHA 14420  
QY 501 -----V-----501  
Db 14421 LASTRLKPSTESTHWLIQERIKTOPNDTAIVSWDGLTYKELGVLASRLAWKQLGLGVG 14480  
QY 502 -----IPL-----VAL--TEL-----L-----V--L-----K 514  
Db 14481 PESLIPLCPKSTWAVVAMVAIEMAGGAFVPLDPNAPVARLGRGIIETDK 14529

## RESULT 12

Q96RW7 PRELIMINARY; PRT: 5636 AA.  
AC Q96RW7  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE Hemictentin.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Trent J.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF156100; AAK68690.1; --  
DR HSSP; P07996; 1LSL.

DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0005021; F:vascular endothelial growth factor receptor. .; IEA.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
DR InterPro; IPR000152; Asx\_hydroxyl\_S.  
DR InterPro; IPR000875; Cecropin.  
DR InterPro; IPR000742; EGF 2.  
DR InterPro; IPR001881; EGF Ca.  
DR InterPro; IPR006209; EGF like.  
DR InterPro; IPR006605; G2F.  
DR InterPro; IPR009030; Grow\_fac\_recept.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003598; Ig c2.  
DR InterPro; IPR000169; Pept\_cys\_acsite.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR009134; VEGFR.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF07645; EGF\_CA; 8.  
DR Pfam; PF07474; G2F; 1.  
DR Pfam; PF00047; IG; 44.  
DR Pfam; PF00090; TSP 1; 6.  
DR PRINTS; PR01832; VEGFRECEPTOR.  
DR SMART; SM00179; EGF\_CA; 7.  
DR SMART; SM00408; IGC2; 43.  
DR SMART; SM00209; TSP1; 6.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 5.  
DR PROSITE; PS00268; CECROPIN; UNKNOWN\_1.  
DR PROSITE; PS01186; EGF\_2; 3.  
DR PROSITE; PS00026; EGF\_3; 5.  
DR PROSITE; PS01187; EGF\_CA; 8.  
DR PROSITE; PS00835; IG LIKE; 44.  
DR PROSITE; PS00639; THIOLEPROTEASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00092; TSP1; 6.  
DR EGF-like domain.  
KW  
SQ

SEQUENCE 5636 AA; 613673 MW; F000B319CED7B52C CRC64;

Query Match 71.2%; Score 2596.5; DB 2; Length 5636;  
Best Local Similarity 9.7%; Pred. No. 1e-41;  
Matches 429; Conservative 69; Mismatches 13; Indels 3905; Gaps 370;

QY 1 M-S-----I-----SS--D-----EV-----8  
Db 1 MISWEVHTVFLFALLYSLLAQDASPQSEIRAEPPGASTLAFVDFVTGSMYDVLVQVI 60  
QY 9 -----NF-LV-----12  
Db 61 EGASKILETSLKRPKRPLNFALVPFHDPEIGFVTITTDPKKFQYELRELYVQGGDCPE 120  
QY 13 -----Y-----R--Y-----LQ-----17  
Db 121 MSIGAIIKTALEISLPGSFYVFTDARSKDYRLTHEVLQIQKQSQVVFVLTGDCDDRTH 180  
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26  
Db 181 IGYKVEIEIATSSSQVFHLDKKQVNEVLKWVEEAQASKVHLLSTDHLEQAVNTWRIF 240  
QY 27 -----T-----PG-----I-----K 31  
Db 241 DPSLKEVTSLSGSPMIEIRNPLGLIKKGFLHELLNIHNSAKVNVYKEPAGMWTYK 300  
QY 32 --S--H-----I-----S--Q-----36  
Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFKKTVSRPVQGIPTVYLLNTSGISTPARID 360  
QY 37 -----S-----NI-----N-----G-----41  
Db 361 LLELLSISGSSLLKTIPTVKYYPHRYKPGIWNISDFVPPNEAFPLKVTGYDKDYLFQVSS 420  
QY 42 -----A-----L-----VP-----P-----A 47  
Db 421 VSFSSIVDPADPKVTMPKTPGYLQPGQIPCSVDSLLPFTLSFVRNGVTLGVDQYLKESA 480



Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIPIINKGDLWGPGLSPKEVKIKVNNLTLECE 2698  
QY 300 A-----K-----Q-----Q-----303  
Db 2699 AYALPSASLSWYKGGQPLKSDHVNIAANGHTLIQKEAQISDTGRTYCVASNIAGEDEL 2758  
QY 304 F-----P-----F-----H- 307  
Db 2759 FVNIQVPPSFQKLWBEIGNMLDTRNGEAKDVIINNPISLYCETNAAPPPTLTWYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315  
Db 2819 LYSDDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878  
QY 316 -----WO-----SN-----319  
Db 2879 LPEEVTVLNKSALIECLSGSPAPRSMQKQDQPLEDDHKKFLSLGRILQILNTQITD 2938  
QY 320 -----NT-----FAS--C-----S 326  
Db 2939 IGRYVCVAENTAGSAKKYFNALNVHVPVSVIGPKSENLTVVNNFISLTCVSGFPFPPDL 2998  
QY 327 -----T-----D-----MCI-----331  
Db 2999 WLKXQPIKLTNTLIVPGGRTLIQIRAKVSDGGEYTCIAXNAGESKKKFSLTVVVPS 3058  
QY 332 --H-----V--C-----K-----L--GO-----338  
Db 3059 IKDHDSESLVNVNREGTSVLECESNAVPPVITWYKGRMITESTHVEILLADGQMLHI 3118  
QY 339 -----D-----R-----P-IK-----343  
Db 3119 KKAESVDTQYVCRAINVAGRDDKNFNLNVVPPSVIEGPEREVIVETISNPVTLTCDATG 3178  
QY 344 --T-----P-----Q-----G-----H-----T-----349  
Db 3179 IPPPTIWLKYNKRIENSLEVRILSGGSKLQIARSQHSDSGNYTCIASNMEGKRAQY 3238  
QY 350 -----N-E-V-NA-----IKW--D--P-----T-----360  
Db 3239 FLISQVPPSVAGAEIPSDVSVLGENVELVCNANGIPTPLIQWLKQKPIASGETERIRV 3298  
QY 361 --G--NLL-A--S-----C--S-----D-----D-----MTL-----373  
Db 3299 SANGSTLNIYALTSDTGKYTCVATNPAGEDEIRFNALNVVPTIRGNKDEAKLMTYVD 3358  
QY 374 --K-----I-W-----SM-----KQ--D-----NCV-----384  
Db 3359 TSINIECRXTGTPPPQINLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418  
QY 385 -----H-----H-----D-----D-----L--Q-----388  
Db 3419 RAGVDNKHYNLQVAFPPNDMSNGTEBITVLKSGSTSMACITDGTAPSMWLBDGQPLG 3478  
QY 389 -----Q-----Q-----H-N-----KE 393  
Db 3479 LDAHILTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKFKILKVLPPHINGSEHEE 3538  
QY 394 I-----YT-----I--K--W-----N-----399  
Db 3539 ISVIVNPLELTCIASGIPAPKMTWMDGRPLPQTDQVTLGGGEVLRISTAQVEDTGRY 3598  
QY 400 -----SPTG-----P-----GT-----N-----407  
Db 3599 TCLASSPAGDDKEYLVRVHVHPNTIAGTDEPRDITVLRNRQVLTLECKSDAVPPVITWL 3658  
QY 408 N-----P-----N-----ANLM-LAS--A--S--F-----420  
Db 3659 NGERLOATPRVILSGRYLQINNADLGTANTVCVASNIAKTTREFILTWNVPNIK 3718  
QY 421 -----DSTV-----RL-W--D--V-----DR-----G-----I-----433  
Db 3719 GPQSLVILLINKSTVLECIAEGVPTPRI TWKDGAVLGNHARYSILENGFLHIQSAHVTD 3778

QY 434 --C-----I-----H-----TL-----T-----439  
Db 3779 TGRYLCMATNAAGTDRRIDLQVHVPPSIAPGPTNMTVIVNQVTTLACEATGPKPSINW 3838  
QY 440 -K--H-----Q-E-----P-V-----Y-----446  
Db 3839 RKNGLHLLNDQNSYRLLSSGLVLIISPSVDDTATYECTVTNAGAGDKRTVOLTVOVPP 3898  
QY 447 SVA-----F-S-----P-D-----453  
Db 3899 STADEPTDELVTKHAPAVITCTASGVPPFSIHWTKNGIRLLPRGDGYRILSSGAIEILAT 3958  
QY 454 -----GRY-----L-----ASG--S--461  
Db 3959 QLNHAGRYTCVARNAGSAHRHVTLLHVHPPVQPOPSSELHVLNPNILLPCEATGTPSP 4018  
QY 462 F-----D-----KCV-----466  
Db 4019 FITWQKEGINVNTSGRNHVLPSGGIQLISRAVREDAGTYMCVAQNPAAGTALGKIKLVQV 4078  
QY 467 -----H-----I-W-----NT-Q-----472  
Db 4079 PVISPHLKEYVIANDKPITLSCADGLPPDPITWKKOGRVIVSIRQVRLSSGSLQIAF 4138  
QY 473 V-----C-----L-----HY-----477  
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTVHVPPRIIRSTEGHYTVNENSOAILPCVADGPT 4198  
QY 478 -----L-N-----QO-----VL-----L-----N-----485  
Db 4199 PAINMKDNVLLANLLGKYTABPYGELILENVVLEDSGFYCVANNAAGEDHTVSLTVH 4258  
QY 486 -----L-G--R-SIC-----L-YT-----LP-H-----H-LV-----500  
Db 4259 VLPTFTPLPGVSLNKGELRLS-CKATGIPLPLKLTWTFNNNIIPAHFDSVNGHSELVIE 4317  
QY 501 -----V-----I-PL-----V-----V-----505  
Db 4318 RVSKEDSGTYVCTAENSQVFKATGVYVKEPPVFKDGYPSNWIEPLGNAILNCEVKG 4377  
QY 506 --AL-----IE-----L-----L-VL-----K 514  
Db 4378 PPTTIQWNRKGVDBISHRIROGLNGSLAIYGTVNE 4413

## RESULT 13

Q8WZ42

ID Q8WZ42 PRELIMINARY; PRT; 34350 AA.  
AC Q8WZ42;  
DT 01-MAR-2002 (TREMELrel. 20, Created)  
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)  
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)  
DE Titin.  
GN Name=TTN;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20309627; PubMed=10850961;  
RA Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereuse F.,  
RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,  
RA Granzier H., Labeit S.;  
RT "Series of exon-skipping events in the elastic spring region of titin  
as the structural basis for myofibrillar elastic diversity.";  
RL Circ. Res. 86:1114-1121(2000).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21573839; PubMed=11717165;  
RA Bang M.L., Centner T., Fornoff F., Geach A.J., Gotthardt M.,  
RA McNabb M., Witt C.C., Labeit D., Gregorio C.C., Granzier H.,



QY 132 -----N-----G-----E-----N-----G-A----- 138  
Db 5669 SGTTCSTATKAGHNCSGHLTVKEPPYFVERQSDVNPTRVOLKALVGGTAPMTIKW 5728  
QY 139 -----H-----T-I-----A----- 142  
Db 5729 FXDNKELHSGAARVWKDDTSTSLFPAKATDSGTIYI COLSNDVGTATSKATLFVKEPP 5788  
QY 143 -----N-----N-----H----- 145  
Db 5789 QFIKKPSPVLVRNGOSTTFECOITGTPKIRSVWYLDGNEITAIQKHGISFIDGLATFOI 5848  
QY 146 -----T-----D-----MM-EV-D-----G- 153  
Db 5849 SGARVENSCTYVCEARNADAGTASCSTELVKYKEPPTFIRELKPVEVVKYSDVELECEVTGT 5908  
QY 154 -----D-V-----E-I----- 157  
Db 5909 PPFVTLKNNREIRSSKYYTLTDRVSVFNHLHTKCDPSDTGEYQCI VSNBEGSCSCSTR 5968  
QY 158 -----PP-----K-----A-----V- 163  
Db 5969 VALKEPPSPFKKIENTTVLKSSATFQSTVAGSPPISTIWLKDDQILDDEDDNVYISFVDS 6028  
QY 164 V-L-R-----GH-----ES----- 170  
Db 6029 VATLQIRSVDNHSGRYTCQAKNESGVERCYAPLLVQEPAPQIVKAKSVDTVTEKDPMTLE 6088  
QY 171 ----- 171  
Db 6089 CVVAGTPELVKWLKDGKQIVPSRYFSMFENNVA SFRIQSVMKODSGOYTFKVENDFGS 6148  
QY 172 -----V-----F-----I-C-----A-W----- 177  
Db 6149 SSCDAVLRVLDQNI PPSTFKLTQMDXVLGSSIHMECKVSGSLPI SAQWPKDGKEISTSA 6208  
QY 178 -----N-----P-----V----- 180  
Db 6209 KYRLVCHERSVSLVNLEEDTANYTCVSNVAGDDACSGILT VKEPPSFLVKPGRQQA 6268  
QY 181 -----S-----D-----LL-AS----- 186  
Db 6269 IPDSTVEFKAILKGT PPKIKWFKDDVELVSGPKCFI GLEGSTSFNLNLSVDASKTGQYT 6328  
QY 187 -----GS-----GDST-----A-R-IV--N- 197  
Db 6329 CHVTNDVGS DCTTMLLVTEPPKFVKLEASKIVKAGDSRLECKTAGSPEIRVVWFRNE 6388  
QY 198 --L-----S--E-N-ST--SG-----ST----- 207  
Db 6389 HELPASDKYRMTFIDSVAVIQMNNLSTEDSGDFICEAQNAGSTSCSTKVI VKEPPVFS 6448  
QY 208 -----QL-----V-----LR----- 212  
Db 6449 FPPIVETLKNAEVSLECELSGT PPFVWVWKQRQLRSK KYKIASKNFHTSIHILNVD 6508  
QY 213 -----HC----- 214  
Db 6509 SDIGEYHCKAQNEVGSDTCVTVLKEPPRFRVSKLNSLT VVAGEPAELQASIEGAQPIFV 6568  
QY 215 -----IRE-----GG----- 219  
Db 6569 QWLKEKEEVIRESENIRITFVENVATLQFAKAE PANAGKYICOIKNDGGMBEENMATLMVL 6628  
QY 220 ----- 219  
Db 6629 EPAVIVEKAGPMTVTGETCTLECKVAGTPELSV EWMYKDKLLITSSQKHKFSFYNKISSL 6688  
QY 220 -----QD-----V-PS-----N----- 225  
Db 6689 RILSVRODAGTYTFQVQNNVGKSCTAVVDVSDRAVP PSTRRLKNKGTGVLGASCILEC 6748  
QY 226 K-----D-V-T-----SLD----- 232

Db 6749 KVAGSSPISVAMFHEKTKIVSGAKYTQTFPSDNNVCT LQLNSLSDSSOMGNYTCVAANVAGSD 6808  
QY 233 -----W----- 233  
Db 6809 ECRAVLTVQEPSPSFVKEPEPLEVLPCKNVTFSTVIRGT PPFKVMFRGARELVKGORCNI 6868  
QY 234 -----N-----S-E-----G-T-L-LL-----A-----T-G- 243  
Db 6869 YFEDITVAELELNFNIDISQSGEYTCVVSNNAGQASCT TRLFVKEPAFLKRLSDHSVPEPGK 6928  
QY 244 -----S-Y-----DGF----- 248  
Db 6929 SIILESTYGTGLPISVTWKDGFNITTSKCNIVTTEKTC ILIILNSTKRDAGQYSCIE 6988  
QY 249 --A-R----- 254  
Db 6989 NEAGRDVCGALVSTLEPPYFVTELEPLEAAVGSVS LQCVAGTPEITVSVWYKGTCLLRP 7048  
QY 255 -----D-G-----N-L--AS----- 260  
Db 7049 TPEYRTYFTNNVATLVFNKVNINDSGEYTC KAENSIGTASSKTVFRIQERQLPPSPARQL 7108  
QY 261 -----TL-----G-----Q-----H 265  
Db 7109 KDIEQTVGLPVTLTCLRLNGSAPIQVCWYRDGVLL RDNLENLOTSFVDNVATLKILQTDLSH 7168  
QY 266 KG-----P-----I-----F-----A--LK 273  
Db 7169 SQOYSCSASNPLGTASSARLTARBP KKSPPFDIKPVSIDVIAGESADFECHVTGAQPMR 7228  
QY 274 --W--NK-----KG----- 278  
Db 7229 ITWSKDNKEIRPGGNYTTCVGNTPHLRLIKVKG DSGOYTCQATNDVCKMCSAQLSVK 7288  
QY 279 ----- 281  
Db 7289 BPPKFVKLEASKVAKQGESIQLECKISGPSPEIKV SFRNDSSELHESWKYNNMFSINSVAL 7348  
QY 282 L-----SA-----GV-D-----K-----T----- 289  
Db 7349 LTINEASAEBSGDYICEAHNGVGDASCSTALT VTKAPPVFTQKPSV GALKGSDVILQCEI 7408  
QY 290 --T--IIW-----D-----A-----H----- 296  
Db 7409 SGT PPEVWVKDRQVRNSKKFKITSKHFDTS LHLNLEASDVGEYHCKATNEVGS DTC 7468  
QY 297 -----T-G--E-A-----K----- 301  
Db 7469 SCSVKFKEPPRPVKLSDTSTLIGDAVELRAIVEGF QPISVVWLKORGEVIRESENTRIS 7528  
QY 302 -----Q----- 303  
Db 7529 FIDNIATLQLGSPEASNSGKYICOIKNDAGMECSA VLTLEPARIIEKPEMTVTGTGNP 7588  
QY 304 F-----P-----F-----H-S--A-----P-A-----L----- 312  
Db 7589 FALECVVCTGTPELSAKWFKDGRELSADSKHHT FINKVASLKI PCAEKSDKGLYSPEVK 7648  
QY 313 -----D-V-----D-----W----- 316  
Db 7649 SVGKSNCTVSVHVS DRI VPPSPFIRKLKDVNAIL GASVLECRVSGSAPISVGVFQDGN EI 7708  
QY 317 -----QS-----N----- 320  
Db 7709 VSGPKQCSFSFBNVCTLNLSLEPSDTGIYTCVAANV AGSDSCSAVLTVQEPSPFEQTPD 7768  
QY 321 -----TF----- 322  
Db 7769 SVEVLPGHSLITTSVIRGT PPFKVMFKGSR ELVPGESCNISLEDFVTELELFEVQPLES 7828  
QY 323 -----ASC-----S-----T----- 327

Db 7829 GDYSLVNDAGSASCTTHLFVKEPATFVKRLADFVETGSPVILEATYTGTPPSVSWI 7888  
QY 328 -D-----M-----C-----330  
Db 7889 KDEYLISQSRCSITMTKSTILEBLSSTIEDYQAQYSLIENAGQDICEALVSVLEPPY 7948  
QY 331 -I-----HV-----CKL-G-----Q-----338  
Db 7949 FIEPLEHVEAIGEPATLQCKVDGTPAIRISWYKEHTKLSAPAYKQFKNNVASLVINK 8008  
QY 339 -----D-----R-----P-----I-KT-----F-----345  
Db 8009 VDHSVGEYCKADNSGVAVASAVLVIKERKLPFPFARKLDVHETLGLFPVAFECRNG 8068  
QY 346 -Q-----G-----H-----T-----N-----350  
Db 8069 SEPLQVSWYKDGVLKDDANLQTSFVHNATLQILQDQSHIGQYNCSASNPGLTASSA 8128  
QY 351 -----EV-----NA-IK--W--D-----PTGN-----362  
Db 8129 KLILSEHEVPPFPDLKPVSDALGESGTFKCHVTGTAPIKITWAKDNREIRFGNYKWT 8188  
QY 363 -----L--L-----A--SC--S-----D-----DM 371  
Db 8189 LVENTATLTVLVKVGKGAGQYTCVASNIAGKSCSAHLGVQEPFRFIKLEPSRIVKQDE 8248  
QY 372 -T-L--KI-----W-----SMK-Q-----D-----C--383  
Db 8249 FTRYECKIGSPEIKVLWYKDET-ETQESSKFRMSFVDSVAVLEMHNLNLSVEDSGDYTCEA 8307  
QY 384 -----V-----H-----D-----LQ-----Q 389  
Db 8308 HNAAGSASSSTSLKVKEPPIFRKKPHIETLKGADVHLECELQGTTPPHVSWYKDKREL 8367  
QY 390 -----H-----N-----391  
Db 8368 SGKKYKIMSENFILSIHLNVDAADIGEYCKATNDVGSCTCVGSIALKAPPRFVKKLS 8427  
QY 392 -----KEI-Y--TI-----K-----W-----399  
Db 8428 ISTVVGKEVOLQTTIEGAEPISVVMFKDGEIVRESNHWISYSENIATLQFSRVEPANA 8487  
QY 400 -----S-----P-----TG-----P-----404  
Db 8488 GKTCQIKNDAGMQCFATLSVLEPATIVEKPSIKVTGDTCTLECTVAGTSELSTKWF 8547  
QY 405 -G--T-----N--P-----N-----A-----411  
Db 8548 KDGKELTSNDKYKISFFNKVSGLKIINVAPSDSGVYSFEVQNPVGKDSCTASLQVSDRTV 8607  
QY 412 -----N-L--L-----M-----L-----ASAS-F-----420  
Db 8608 PPSFTKLTNGLSGSSVVMCKYGVSPISVSWFHEGNEISSGRKYQTLTDNTCAL 8667  
QY 421 -----DS-----TVR-----L-----426  
Db 8668 VNMLESDSGDYTCIATNAGSDECSAPLTVPSPSVQKPDMDVLGTGNTVFTSIVKG 8727  
QY 427 -----W-----DVD--R-G--ICI-----435  
Db 8728 TPFPSVSWFKGSELVPGDRCNVSLSDSVAELDFVDTSQSGEYTCIVSNEAGKACTT 8787  
QY 436 H-----TLT-----K-----H-Q-----442  
Db 8788 HLYIKAPAKVKNLDYSIEKGPLILEGTFPTTPISTVTKNGINVTSPQRCNITTTTE 8847  
QY 443 -----E-P--V-----YS--V-----A-----F-----SP-----D-----453  
Db 8848 KSAILEIPSTVEDAGQYNCYNIENASGKSCSAQILILEPPYFVKQLPEPVKVSVDASIL 8907  
QY 454 -----G-----R-----Y-----L-----A--S-G 460  
Db 8908 QCQLAGTPEIGVSWYKGDTKLRPTTTYTKMFRNNVATLVFNQVDINDSGEYICKAENS 8967

QY 461 -----SF-----D--K-----C-V-----HI--W-----469  
Db 8968 EVSASTFLTVQEOKLPSPFSRQLRDQVETVGLPVVDFCAISGSEPIVSWYKDKPLKDS 9027  
QY 470 -----NT-----QV-C-----L-475  
Db 9028 PNVQTSFLDNTATLNIFKTRSLAGQVSCATNPIGSASSARLLILTEGKNPPFFDIRLA 9087  
QY 476 -----H-----Y-L-N-----479  
Db 9088 PVDVVGESADFECHVTGTQPIKVSNAKDSREIRSGGKYQISYLENSAHLTVLKVDKGDS 9147  
QY 480 GO-----V-----L-----N--L-GR--S-----489  
Db 9148 GQYTCYANVEVGKDSCTAQINIKERLIPSFYKRLSETVEETEGNFKLEGRVAGSQPIT 9207  
QY 490 -----I-----C-----LYT-----L-----495  
Db 9208 VAWYKNNIEIQPTSNCEITFKNTLVQVRKAGMNDAGLYTCKVSNDSAGSALCTSSIVIK 9267  
QY 496 -----P-----H-----498  
Db 9268 EPKPEVFDQHLTPVTVSEGYVQLSCHVQSGSEPIRIOWLKAGREIKPSDRCSFSPASGT 9327  
QY 499 -----L-----499  
Db 9328 AVLELRDVAKADSGDYVCKASNVAGSDTTKSKVTIKDKPAVAPATKAAVDRGLFFVSEP 9387  
QY 500 -----VV-----IP-----503  
Db 9388 QSIIRVVEKTTATFIKVGDPPIPNVKWTKGKWRQLNQGGRVFIHQKGDRAKLEIRDTTKT 9447  
QY 504 -----L--VA-----L-----IE-L--L-----VLK 514  
Db 9448 DSGLYRCVAFNEHGEIESNVNLQVDERKQEKIEGDLRAMLKKTPIK 9495

## RESULT 14

Q6XA09 PRELIMINARY; PRT; 7191 AA.  
AC Q6XA09  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nonribosomal peptide synthase.  
GN Name=NRPS1;  
OS Alternaria brassicae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;  
OC Pleosporales; Pleosporaceae; Pleosporaceae; Pleosporaceae; Alternaria.  
OX NCBI\_TaxID=29911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14727058;  
RA Guillemette T., Sellam A., Simoneau P.;  
RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria  
brassicae and flanking genomic sequences.";  
RL Curr. Genet. 45:214-224 (2004).  
CC -!- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme  
family.  
CC EMBL; AY246697; AAP78735.1; -;  
DR InterPro; IPR010071; AA\_adenyl\_dom.  
DR InterPro; IPR009081; ACP\_like.  
DR InterPro; IPR000873; AMP\_bind.  
DR InterPro; IPR001242; Condensatn.  
DR InterPro; IPR000977; DNA\_ligase.  
DR InterPro; IPR006162; Ppantne S.  
DR InterPro; IPR006163; Pp\_bind.  
DR Pfam; PF00501; AMP-binding; 4.  
DR Pfam; PF00668; Condensation; 9.  
DR Pfam; PF00550; PP-binding; 6.  
DR PRINTS; PRO0154; AMPBINDING.  
DR TIGRfams; TIGR01733; AA-adenyl-dom; 4.

DR PROSITE; PS00075; ACP DOMAIN; 7.  
DR PROSITE; PS00455; AMP BINDING; 2.  
DR PROSITE; PS00697; DNA LIGASE A1; UNKNOWN 1.  
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN 3.  
SQ SEQUENCE 7191 AA; 793216 MW; F908905155F2394C CRC64;

Query Match 71.0%; Score 2590.7; DB 2; Length 7191;  
Best Local Similarity 9.4%; Pred. No. 2.2e-41;  
Matches 442; Conservative 55; Mismatches 9; Indels 4203; Gaps 363;

QY 1 M-----S-----IS-----S-----D-----E----- 7  
DB 1521 MAFRLAGSGRAPAGHAETTLARLWASVLKLDAGAVGAEDSFRLGGDSISAMKLVTAAR 1580  
QY 8 -V-N---F---L---V---V---Y---R----- 14  
DB 1581 KGGVNLVASVFAQPKLLMAATAVAVMLPSDDAAKPEADTLPMELLPASSRQAVALAASE 1640  
QY 15 -----Y-----LOE----- 18  
DB 1641 CDVFPDCIEDMYPCKLQEGVLMTNKPCTYVVOPIYRLPADMDISRFEAKVVAEE 1700  
QY 19 -----S---GF-----S---H---S-----A-FT----- 27  
DB 1701 ASLRTIVYSEBHGFLQVVVREDIKQWLPDIQHINETTRLPKANGAPLATFTLVGENT 1760  
QY 28 -F-----G-----IK---S-H-----IS---QS-- 37  
DB 1761 DSMFFWTAHVAVDGWSWTALPRKVEAYRSGVHSMPATVPYSRVKYISSLDORQSDA 1820  
QY 38 -----NI-----NGAL---V---P-----P---AA---L--- 49  
DB 1821 FWLTQLDNVTAQFPQLPSPDRHVEANGQLLTVLTRNPGLEVTVPVMIRAAWGILLAT 1880  
QY 50 -----I-SII-----OK----- 55  
DB 1881 YSGSDVIMGETNSGREASVPGIESITPTTAPVRLRLNRLSLTVHLYKETOKQSSLS 1940  
QY 56 -----GLO-----Y-VE---AE---VS---I----- 66  
DB 1941 LPYQFAGLOHISKLSPEATAICDFQSLGIEAGDDFDAESSLWNWVSANTIQTDFSYAF 2000  
QY 67 -----N-----E-----D-----GT----- 71  
DB 2001 VFNCNVATGVQVEALFDDRVRVQTWLAQRVQVQFDFILQLNGTDNLRLSLDLDLINEA 2060  
QY 72 -----L-----F-----D----- 74  
DB 2061 DRKTSSMNSKEVPPIIPRCHSVIAEDOTALRPTACAIADWDGTGMSYRELDERSALAH 2120  
QY 75 -----G-RP-----I----- 78  
DB 2121 QLIRLGRPKQVPLCFDKSGMTVAIIAVLKAGAAFPVLPDPEAPVLRUREIVSDIDL 2180  
QY 79 -----E-----S-----L-SL----- 83  
DB 2181 LLCAPQYRELQOSIPCTSTWVDRQATETVAGRLPSLPSVHSDSPAYAFYTSSTGKPKCA 2240  
QY 84 -----I-----DA---V---M-----PD----- 90  
DB 2241 VVHHTWVTSSTAFAAGWKISTASRVLPQFASYTFDACLIEVFSTLMQGGTVCVPOGSR 2300  
QY 91 -----VV-----Q----- 93  
DB 2301 NDLVGVINRFNVNWAALTPSVVRMIVPSEIPOLETFLVGEAMSQDVLVTWADKNVNG 2360  
QY 94 -----T-----R----- 95  
DB 2361 YGPTCAAVATSNIMTPHMRPNLGRAVTARGWIVSRNNHTLAPVGAIGELLLEGGAVG 2420  
QY 96 -----Q-----QA-----Y-R-D----- 101  
DB 2421 AGYLNPNKTAQVFGQARWCVGLMGDDISAPVRIYKTDGLVKYNEGDGTMVYLRKDLQ 2480

QY 102 -----KL-----A-----Q----- 106  
DB 2481 KVRGQRLSELSEVEHKLDDHVMVQSALASVPTTGPCAKRLVAIVSLQKHGHDHDTDDKQLR 2540  
QY 107 -----Q-A---A-----A-----A-----A-----A 113  
DB 2541 LLPQENASLINIATIRDCGLCERLPAYMIPSLMTIAVERFPLMPGSKMDRRCAIQWLEQMDQA 2600  
QY 114 -----A-----A-----A-----A-----AA----- 118  
DB 2601 TYRLISAMGTDDAEVNNGSPIEKRLRAIPAKVLNLSVADVRLNQSLHGLGDSIAAMQVS 2660  
QY 119 SQ---QG----- 122  
DB 2661 SOCRAQGPISVODIIRSKSISAMASAVDLSOSSQVTTAKDYDLPDPLSPIQKVFDDAV 2720  
QY 123 -----SA-----KN---G-----E----- 128  
DB 2721 GDKHNYFNQTEFLRLSRNIETELRSALTVLAKTHPLRAARFKNAGAKQRIEKDVSS 2780  
QY 129 -----NTAN-----G-----E----- 134  
DB 2781 SPRLRHHVQAGNDANLRPIIDHSQATLDVAKGPTFAIELFDVDDTFSQAIALVAHHLII 2840  
QY 135 -----EN---G---A-----H---T-----IA----- 142  
DB 2841 DVVSGILLEDLQGLQLOPAPQSPMPYHMLHEQSLQATQESARRVPVPGDIAPGDLDY 2900  
QY 143 -----N-N-----H-T-D-MM-----E---V-D----- 152  
DB 2901 WAMEGRPNVNGDVBEEDLHLSTRDTMLLGAQDALATETIDILIAALFESFRKVPFDRST 2960  
QY 153 -----G-----D-----V---E-----I----- 157  
DB 2961 IAHNEGHRETNNRQDLRSRTVGWFSVTVPPIHLFVPLNEATDMISTIRWDRFRNRPD 3020  
QY 158 -----P-----P-----NK----- 161  
DB 3021 KCRPYFAYNLTEEGQTRFASHWPAEAVNYVGRSQDNKQGLFTALSDVDSREVGEV 3080  
QY 162 -----L-----R-----G-----H-E----- 169  
DB 3081 PRALFDITAAVSQGAIKLSFGMNRMRQKEIRAWVGKRCQTLVDVAVELLQARQERSV 3140  
QY 165 -----L-----R-----G-----H-E----- 169  
DB 3141 GNFKYLPPLYNGISRLSAILPAGINLNDVEDIYPASPMQOGLLHTQSRHPELYTYHTVSQ 3200  
QY 170 -----S-----E-----VFI----- 174  
DB 3201 VQSADGNPIDPRLAEMQVWVHRHQALRTIFIDSLAKDGSKDQIVLKEKPGRVQILADC 3260  
QY 175 -----C-A-----W-----N----- 178  
DB 3261 DDSQVANLLRHQSIDCREALPPHRMSICKTKTGRWPKLELHSHVINDGTSVSNLLADLA 3320  
QY 179 -----P-----V-----S-DL-L-A-SG-----S 188  
DB 3321 RAYARKLTRADAGPLYSDIYGYMLSRSSDADLAYKHAHLSGIEPCLPFLVNDGIPSPSPS 3380  
QY 189 G-D---STARI-----W-----N----- 197  
DB 3381 GSDVDELGSTRKQVFCQNGVTLNVLQTLWALTLYHYVGTVDVSFGLIASGRDIPVTN 3440  
QY 198 LSE-----N-----S---T-----SGST-----QL----- 209  
DB 3441 INEAVGCFNMVMSRLSPSDETTIAQLLEALQGTSEALSHQGCSLADIQHALQPLSLFN 3500  
QY 210 -----V-----L-----R----- 212  
DB 3501 TAFTFQRSLSDPBDTALYEDMEADAGEYIVTVNADVTQOSITVDFGYSKDRILPSQ 3560

213 QY -----H-----C-----I-- 215  
3561 DB AONMAETFKILDIVVCSASELTIGKLDILTTESSIHQIMENWPOLPPPIRCLHDVIHD 3620  
216 QY -----R-----EG--G--QD-----VP----- 223  
3621 DB QALTRPRTTKAVEGWDGTFYQDFDKITNQLAVHLQSIGVTTFTFVILPEKSYAIVSM 3680  
224 QY ----- 223  
3681 DB IAIMKAGAVVPLDPKHPQTRLELIEDVGASVVLCSRGYHTTIASEVAKTAVIVDQRSR 3740  
224 QY -----S-----N-----K----- 226  
3741 DB KLGVPISSPKRTCATPDNAYCLFTSGTGKPKGTIIPHQAFCTSAAAFTRMNINATSR 3800  
227 QY -----D-----VTS--L-----D----- 232  
3801 DB TFQFASYTFDASCIEILSALTVGATVCVPTEDDRMNAAGAIRKLRVNMSELLTPSVLGTI 3860  
233 QY -----W--NS-----E----- 236  
3861 DB EPERVGLKTLVSGEALSGPILKKWSNSTCFINAYGPTSCSVAAATAYKSTLDHKLIVS 3920  
237 QY -----Y-----D-----G--FAR-----I-- 251  
3921 DB EPCT--IGTSGCRLWIVHPRNHDKLMVPGSVGELVIEGPTVARGYLNDVKTAKAFINDP 3979  
252 QY -----T-----K-----DG--NL--AST-----L--GO-----H-- 265  
3980 DB AWAKTIFSSNNTFEAARMYKTDLVRYNTDGSVNYIGRKDTQIKLANGRIELGEIBFHV 4039  
266 QY ----- 265  
4040 DB KNPPERVOQSAVELVAPNSRSSAKALAVFAVVOQDAIDGEOQSVQPASTDLPAAADLLLLPL 4099  
266 QY -----K-----G-----P--IF--ALKW----- 274  
4100 DB SDELDMCKNTENGLAGSLPSYMPIAIFIPVTKLPWTSAGKLDNRNLSIVQLNSRETMA 4159  
275 QY -----NKK-----G----- 278  
4160 DB MYRLTSIANKKPKITEAEKIKHKAVCVSLPSPSSVGIDDSFVRLGGDSISSMRLVAMAH 4219  
279 QY -----N-----F-----I----- 281  
4220 DB TEQWELSFIDIFKNPKSLDLAKIGAQISKSQAQKVMQPPDLLPASLITRSDVISEVVQOC 4279  
282 QY -----LS--AG--V-----DKT----- 289  
4280 DB QVSKEDLDAYPTSSQLDALLTSLIKQAGAYVAQHVLAALPKSLDMTKFKAAWESAIOEID 4339  
290 QY -----T--II-----W-----DA-----H-----T-- 297  
4340 DB ILTRTIQMPGIFMOAVLNDPVDREAKSLKSAEDDASKIPPHLGHLAAVTLVTPTS 4399  
298 QY -----A-----K-----Q----- 302  
4400 DB GERYFVWTLHHALYDGSYILMLQVQOQIYKGVSTTPQTSYARFVEYLSSTSVSDSDVY 4459  
303 QY -----QFP-----F--HS--A-----PA----- 311  
4460 DB WRERLTGVNAYQFPRPSHATSSAPPNGQMFQHSKIAHRKNTDVTTPANAIKAAWALLIAA 4519  
312 QY -----L-----DV----- 314  
4520 DB YTGSDDVVFGETLAGRDVAMTGITDVCGPTLTTPVSRVKIDRGATVSDLLNTIATNITDR 4579  
315 QY -----D--W-----Q----- 317  
4580 DB IPYQHHGLSAIKALGEDMIAACDFQNLVLOTENEELADSMWSVHDBEQGNFTFYPLVI 4639  
318 QY -----S-----N----- 319

4640 DB ECKWGLSKTEVLAHFPAANVISLWHVORLIYQFETVLIQOSATHVRIHIAVLSDQDKQVLR 4699  
320 QY -----N-----T-----F-----AS----- 324  
4700 DB KWNAYEPRLIDDTVPSPFLFKAAASQPTTTAVTAFDGEFSYVGLSALASQALQELVKFGAG 4759  
325 QY -----C--S-----TD----- 328  
4760 DB PECLIPICVDKGRMAIVAILAILISGAGVYVPLSPDTPASRHLHI VETKASIVLCSPKYT 4819  
329 QY -----M-----C-----I--H-- 332  
4820 DB HRFVEMVGHVPSVSETAIRQLPTSSISLSQRAKNNICYVITFSGSTGLPKGVVIEHKS 4879  
333 QY -----VKLG-----Q--D-----R-----P----- 341  
4880 DB VSSAAICE--GLHITPTSRVQFCSFLFDVSVGETLTVLLRGATICVPSDEORTTLNAAA 4938  
342 QY IKT-----F-----QG-----HTN----- 350  
4939 DB V-TDLNANWAFITPVSASTLEGPKSVPTTLETLVGGEAMTSDVVDKMATGVNLH--NGYGP 4996  
351 QY --E--V--N--A--I--K--W--D--P--T--G-- 361  
4997 DB TEGTVFAIGNDHVSAQRDPNSIGHPLKSGRAMLTNSDNPHELAPIGATAELCLEGPLLAR 5056  
362 QY -----N-----L--LA--S----- 366  
5057 DB GYLNDPKRTSEAFIEAPAFKFNFSNSSESIYRTGDLVQYAADGSIQYMGKRDNIQKLAG 5116  
367 QY -----C-----SD----- 369  
5117 DB QRIELDEIOAVHANNVHVQVQPKVCPCKKLTVVVSPFGTAASNAGSWRRILSDT 5176  
370 QY -----D-----M-----TL-----K--I--W--SM--K--Q-- 380  
5177 DB ESLSQINRARDRLADLVPSYVMPFIWIAVPRIPTLASAKLDKKQVGLWLEGMDEALYQRI 5236  
381 QY -----D-----N-----C----- 383  
5237 DB MGAELPEDMEGPGAAALTVLIRGIWAKVLRNRPVEDVKPSKWSLSLGGDSISAMKLLAKCRS 5296  
384 QY -----V-----H--D--L--O-----Q-----H 390  
5297 DB EGINLNLQILRAKSLSHLAADVKSVMILDEKEQNDRPFPALSPIQRFVFEAGSIENSTH 5356  
391 QY -NKE-----I--Y-----TI-----K-----W-----SP----- 401  
5357 DB FN--QSSTLRILRYVQPMVQORALNSIVECHSMRLARFSDKNNGQWQOLVMSKVSGSYAFT 5415  
402 QY -----TCP----- 404  
5416 DB AHDVSTASAAGAIISSTOKSLDIRTPGFVAVDLFNLKGHQLMFLMAAHLHVIDVWSGILL 5475  
405 QY -----G-----T--N-----N----- 408  
5476 DB GLEDLGSGPVTTLPRSLPFTQWCKMQTSSASEITQOLTVKXQPLVVEPANFAFWGMDV 5535  
409 QY -PN-----A--N-----L--MLASA--SFDS--TV 424  
5536 DB RPNVYGDSEDRDEFVIDKETSAMAFDNHHVYKTDLDVILIA--AILHSF--SRVFINRKAPTL 5593  
425 QY -----R--LWD--VD--R--G-----I-----C--IHTL-----T--K----- 440  
5594 DB FNESHGREVMDGSLDLKRTVGWFTTLYPVTVPIDDDEDEVHTLRQVKDTRRKVASNR 5653  
441 QY -----HO-----E-----P--V--Y-----SVAFSP-----D-- 453  
5654 DB PYFAHRYLTEDGKERFANHAPMEVLFNVLGRQEQSGHSDSL--LSPQVEGDDDETSVGV 5712  
454 QY -----G-----RY-----LA----- 458



```

Db 5713 KTSRMALFEISAVTEGQIQLSFMVNYRKNKGIHRTWTAECORTLEETIAELAKTRDQ 5772
QY 459 -----S-----G-SFD-----KC-V--HI-----W 469
Db 5773 PTWADPPLPLBSYSRLDLVLTPLHAGVPSFDQVEDMYPCSPIQDMILSQIKSPESYW 5832
QY 470 --NT-----QV-----473
Db 5833 SSTTEVRSKRGPVDASKVVDGKQVVARHPALRTIFDSVCKGVDFDQIVVKSPDSGIV 5892
QY 474 -----C-----L--H-YLANG-----Q-----VLN-----485
Db 5893 TYKDAEALATLESIRHSSSLGKKPVLPHQAUVVQTSKGKIFVKVIVNHAVIDGGSIG 5952
QY 486 -----L-----GR-----S-----I-----C-----491
Db 5953 VIGQDLQEAEGRLSDGLYSAYIKYLRALPAEDAIAWYKAKLRGVSCYFPTTPRDP 6012
QY 492 -----L-----L-----YT-----494
Db 6013 KPRQLQSLDMRFRFDELHDLAESSNVTTIANLLAAWALVLRSYTNSDVICYLTSGRN 6072
QY 495 -----I-----I-----PH-----H-----498
Db 6073 VPIDRIENAVGAFINMLVSRIELKSATSLLEIVENVQSDVFGVSMPHQHCSLAQFQHD 6132
QY 499 -----L--VV-I-----P--L-----V-----AL--I-----508
Db 6133 SKSLSFNTAVSIQRSSPELSDSGIEFQDLGDHPSEFAITVVIDATRDGVRFTY 6192
QY 509 -----E-----L-----L-----VL-----K 514
Db 6193 WSDAVTGDGEAKNVSTLMKILVQVLSNPK 6221

RESULT 15
AAP78735
ID AAP78735 PRELIMINARY; PRT; 7191 AA.
AC AAP78735;
DT 01-APR-2004 (TREMELrel. 27, Created)
DT 01-APR-2004 (TREMELrel. 27, Last sequence update)
DE 01-APR-2004 (TREMELrel. 27, Last annotation update)
DE Nonribosomal peptide synthase.
GN NRPS1.
OS Alternaria brassicae.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.
OX NCBI_TaxID=29911;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=14727058;
RA Guillemette T., Sellam A., Simoneau P.;
RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria
RL brassicae and flanking genomic sequences.";
RL Curr. Genet. 45:214-224(2004).
DR EMBL; AY246697; AAP78735.1; .
SQ SEQUENCE 7191 AA; 793216 MW; F908905155F2394C CRC64;

Query Match 71.0%; Score 2590.7; DB 2; Length 7191;
Best Local Similarity 9.4%; Pred. No. 2.2e-41;
Matches 442; Conservative 55; Mismatches 9; Indels 4203; Gaps 363;

QY 1 M-----S-----IS-----S-----D-----E-----7
Db 1521 MAFRLAGSGRAPAGHAEVTLARLNASVLKLDAGAVGSDSPFRLGDSISAMKLVTAR 1580
QY 8 -----V-N-----F-----L-----V-----Y-----R-----14
Db 1581 KGVVLNVASVFAQPKLLAMAATAVMLPSDDAKPEADTLPMELLPASSRQAIVALAASE 1640
QY 15 -----Y-----LQ-----18
Db 1641 CDVFPDCIEDMYPCSKLQBLVMTLNKDPGTYYVQPIYRLPADMDISRPFKAWKVVAE 1700

```

```

QY 19 -----S--GF-----S-----H--S-----A-FT-----27
Db 1701 ASLRTRIVYSEEHGFLQVVVREDIKWQSLDPIQHINETTRRLPAKNGAPLATFTLVGENT 1760
QY 28 -----F-----G-----IK-----S--H-----IS-----QS--37
Db 1761 DSMFFVWTAHHAHVDCWSWTALFRKYANRSGVHSMPTATPYPSRFVKYISSLDQSDA 1820
QY 38 -----NT-----NGAL--V-----P-----P-----AA--L--49
Db 1821 FWLTQLDNVTAAQFPQLSPDRHVEANGOLLHTVLLTRNPGLEVTVPVSMIRAAWGILLAT 1880
QY 50 -----I-SII-----55
Db 1881 YSGSDDVIWGETNSGREASVPGIESIIGPTITTAAPVRLNRSLTVHDYKKEQKQSSLS 1940
QY 56 -----GLQ-----Y-VE-----AE-----VS--I-----66
Db 1941 LPYQFAGLQHISKLSPETAIACDFQSLGIEAGDDFDESSLMNMVMSANTIGTDFPSYAF 2000
QY 67 -----N-----E-----D-----GT-----71
Db 2001 VFNCKVATGVOVEALFDDRVRVQVTLAORMVOQDFILTQLNGTNDILSLDLDLINPA 2060
QY 72 -----L-----L-----F-----D-----74
Db 2061 DRKTISSMNSKVPPIIPRCHSVIAEDQTALPTACIDAWDTGFMYSRELDERSALAH 2120
QY 75 -----G-RP-----78
Db 2121 QLIRLGVRPKQFVPLCFKDSKGMTIVAIIVAKAGAAVPLDFEAPVLRUREIVSDADL 2180
QY 79 -----E-----S-----L-SL-----83
Db 2181 LLCAFOYRELQOSIFCSTWVVDROATETVAGRLPSLPSVHSDSPAYAFYTSGSTGPKGA 2240
QY 84 -----I-----I-----DA-----V-----M-----PD-----90
Db 2241 VVHHTHWTSSTAPAGWKISTASRVLOFASYTFDACLLEVEFTLMQGGTVCVPDQGSRT 2300
QY 91 -----VV-----Q-----93
Db 2301 NDLVGVINRNVNWAALTPTSVVRMIVPSEIPQLETLFLVGEAMSQODLVTWADKVNLGNG 2360
QY 94 -----T-----R-----95
Db 2361 YGPTCAAVATSNIMTPHMRPNNLGRAVTARGMIVSRNNHHTLAPVGAIGELLLEGAVG 2420
QY 96 -----Q-----Q-----Y--R--D-----101
Db 2421 AGYLNNPETAQVVFVQARWCVGLMGDDISAPVRIYKTGDLVKYNEGTMLYLGRKDLQT 2480
QY 102 -----KL-----A-----Q-----Q--106
Db 2481 KVRGORLELSEVEHKLDDHMQSALASVPTTGPCKAKRLVAIVSLQKHGDHDTDTDKQLR 2540
QY 107 -----Q--A-----A-----A-----A-----A 113
Db 2541 LLPOENASLNIAITIRDGLCERLPAYMIFSLMTAVERFPLMPSGKMDRRCAIQWLEQMDQA 2600
QY 114 -----A-----A-----A-----AA-----118
Db 2601 TYRLISAMGTTDAEVNNGSPIEKKLRAPAKVNLNSVADVRLNQSLHGLGSDSIAMQVS 2660
QY 119 SQ--QG-----122
Db 2661 SQCRAQGPISVQDIIRSKSISAMASAVDLSOSSQVTTTEAKDYDLFPDLSPIQKVFFDAV 2720
QY 123 -----SA-----KN--G-----E-----128
Db 2721 GDKHNYFNQTEFLRLSRNIETBELRSALTVLAKTHPMLRARFSPKNEAGAKORIEKDVSS 2780

```

QY 129 -----NTAN-----G-----E----- 134  
Db 2781 SFLRHHVQAGNDANLRPIIDHSQATLDVAKGPTFAIELFDVDDTFPSQAIALVAHLLII 2840  
QY 135 -----EN-----G--A-----H-----T-----IA----- 142  
Db 2841 DVVSWGILLLEDQLGLQLPAPQSMYPYHMLHEQSLQATQESARRVFPVGDIPAGDLDY 2900  
QY 143 -----N--N-----H--T--D--MM-----E-----V--D----- 152  
Db 2901 WAMEGRPNVNGDVVEEDLHLSTRDTWLLGQAODALATETIDILIAALFESFRKVPFDRST 2960  
QY 153 -----G-----D-----V-----E-----I----- 157  
Db 2961 IAHNEGHEGTNNRQDLSRTVGFSTVTPHLPVPLNEATDMISTIRWRDFRNRTPD 3020  
QY 158 -----P-----P-----NK----- 161  
Db 3021 KGRPFAYRNLTEEGQTRFASHWPAEAVFNYYVGRLOQDNKDGCLFTALSDVDSREVGEDV 3080  
QY 162 -----C--A-----W-----N----- 178  
Db 3081 PRLAFDITAAVSOQAISLFGWNRMRQKEIRAWVGKRCQOTLVDVABEELLOARQERSV 3140  
QY 165 -----L-----R-----G-----H--E----- 169  
Db 3141 GNFYLPPLLYNGISRLSAILPAGINLNDVEDIYPASPMQOGLLHTQSRHPELYTYHTVSQ 3200  
QY 170 --S-----E-----VFI----- 174  
Db 3201 VQSADGNPIDPRLAEAWQVVHHRHQALRTIFIDSLAKDQSKQIVLKEKPGRVQILADC 3260  
QY 175 -----C--A-----W-----N----- 178  
Db 3261 DDSQVANLLRHOSSIDCREALPPHRMSICKTKGRVWFKLELSHVINDGTSVSNLLADLA 3320  
QY 179 -----P-----V-----S--DL--L--A--SG----- 188  
Db 3321 RAYARKLTRADAGPLYSDYIGYMLSRSSDADLAYWKAHLSGIEPCLFPVLNDGIPSSPES 3380  
QY 189 G--D-----START-----W-----N----- 197  
Db 3381 GSVDELGSTRVQDFCKQNGVTLSNVQLTWALTLYHYVGTVDVSFGLIASGRDIPVTN 3440  
QY 198 LSE-----N-----S--T-----SGST-----OL----- 209  
Db 3441 INEAVGCFVMVSRLSFSDETTIAQLLEALQTSREALSHQCGLADIQHALLQPSLFN 3500  
QY 210 -----L-----R----- 212  
Db 3501 TAFTRFORSLSDPEDTALIYEDMEADAGEYIVTVNADVTQDSITVDFGYSKDRILPSQ 3560  
QY 213 -----H-----C-----I-- 215  
Db 3561 AQONMAETFKILDSIVVCSASELTIGKLDILTESSIHQIMWNPQLPPPIRRCLHDVHD 3620  
QY 216 -----R-----BG--G-----QD-----VP----- 223  
Db 3621 QALTRPRTTKAVERGWDGTFYQDFDKITNQLAVHQSIGVTTTFVPIILFEKSSYAIVSM 3680  
QY 224 ----- 223  
Db 3681 IAIMKAGGAYVLPDPKHPQRLRELIEDVGASVVLCSRGVHTTASEVAKTAVIVDORSIR 3740  
QY 224 -----S-----N-----K----- 226  
Db 3741 KLGVPISSEKPTCATPDNAAAYCLFTSGTGKPKGTIIIPHOAFCTSAATFRMNINATSR 3800  
QY 227 -----D-----VTS--L-----D----- 232  
Db 3801 TFQFASYTFDASCIEILSALTVGATVCVPTEDDRMNNAAGAIRKLNRWMSLLTPPSVLGTI 3860  
QY 233 -----W--NS-----E----- 236

Db 3861 EPERVPLKTLVSGGEALSGPILKKWSNSTCFINAYGPTFECSSVAAATAYKSTLDHKLIVS 3920  
QY 237 --GTLIATGS--Y-----D-----G--FAR-----I----- 251  
Db 3921 EPGT--IGTSGCRLMTVHPRNHDKLPVGSVGVELVIEGPTVARGYLNDVKTAKAFINDP 3979  
QY 252 -W-----T-----K-----DG--NL--AST--L--GO-----H-- 265  
Db 3980 AWAKTIFSSNNTFEARMYKTDGLVRYNTDGSVNYIGRKDTQIKLNGORIELGEIEFHV 4039  
QY 266 ----- 265  
Db 4040 KNFPERVQSAVELVAPNRSSEKALAVFAVVDQDAIDGEOQSVQAPASTDLPAAADLLPL 4099  
QY 266 -----K-----G-----P--IF--ALKW----- 274  
Db 4100 SDELDMCKNTENGLAGSLPSYMIPIAIFIPVTKLPMTSAGKLDNRNRLSLVQNLSETWA 4159  
QY 275 -----NKK-----G----- 278  
Db 4160 MYRLTSIANKKPIITEAEKIIHKAVCVLSLPPSSVGIDDSFVRLGGDSISSMRLVAMAH 4219  
QY 279 -----N-----F-----I----- 281  
Db 4220 TEQMELSFIDIENPKLSDLAKIGAKISKOAEKVMQFPDLLPASLTRSDVISEVVQOC 4279  
QY 282 -----LS--AG--V-----DKT----- 289  
Db 4280 QVSKEDQDAYPTSSIQDALLTSLIKQAGVVAQHVLAHPKSLDMTKFKAWEAISQIEID 4339  
QY 290 --T--II-----W-----DA-----H-----T-- 297  
Db 4340 ILRTRIOMPSGIFMQAVLRNDPVDWREAKSLKSAEDDASKIPPHLGHLAAYTLTTPS 4399  
QY 298 GE-----A-----K-----Q----- 302  
Db 4400 GERYFWTLLHALYDGMISYLMQRVQIYSGVSTTPQTSYARFVEYLSSTSVSDSVY 4459  
QY 303 -----QPP-----F--HS--A-----PA----- 311  
Db 4460 WRERLTGVNAYQFPRPSHATSSAPPNGOMFOHSMKIAHRKNTDVTAPANAIRAAWALILAA 4519  
QY 312 --L-----DV----- 314  
Db 4520 YTGSDDVVEGTLAGRDVAMTGITVCGFTLTVPSRVKIDRGATVSDLLNTIATNIDR 4579  
QY 315 -----D--W-----Q----- 317  
Db 4580 IPYQHGLSAIKALGEDMIAACDFQNLVLIQTENEELADSMWSVHDNEEQGNFTYPLVI 4639  
QY 318 -----S-----N----- 319  
Db 4640 ECKWGLSKTEVLAHFADANVISLHWQRLIYQFETVLILQSAATHVRHIAVLSDQDKQLVR 4699  
QY 320 --N-----T-----F-----AS----- 324  
Db 4700 KMWAVEPRLIDTVPSELFPFKKAASOPTTAVTAFDGEFSGYSELASALQAEVLKFGAG 4759  
QY 325 -----C--S-----TD----- 328  
Db 4760 PECLIPICVDKSRWAIVAILAILISGAGVYPLSPDTPASRHLHI VETCKASIVLCSPKYT 4819  
QY 329 -----M-----C-----I--H----- 332  
Db 4820 HRFVEMVGHVFSVETAIRQLPTSSISLSORAKNNICVIFTSGSTGLPKGVIEHKSI 4879  
QY 333 -----VCKLG-----Q-----D-----R-----P----- 341  
Db 4880 VSSSAAICE--GLHITPSTRVFCFLFVSGVETLTVLLRGATICVPSDEQRTNLAAA 4938  
QY 342 IKT-----F-----QG-----HTN----- 350

Db 4939 V-TDLNANWAFLETPSVASTLEGKSVPTLETTLVVGGEAMTSDVVDKWTGVNLH-NGYGP 4996  
QY 351 -E-V-N-A-I-K-W-D-T-P-T-G- 361  
Db 4997 TEGTVFAIGNDHVSAQDPDSNCHPLKSGRAMLTNSDNPHELAPICATAECLCEGLPLAR 5056  
QY 362 -N-LA-S- 366  
Db 5057 GYLNDPKRTSEAFIEAPAFKPNFSSNESRIYRTGDLVQYAADSGIQYMGKDNQIKLAG 5116  
QY 367 -C-SD- 369  
Db 5117 QRIELDEIOAVHADNNHVQVQLPKVGECTKLTVVVSPFGTAASNAGSDRRILSDT 5176  
QY 370 -D-M-TL-K-I-W-SM-K-Q- 380  
Db 5177 ESLSQINRRADRLADLVPSVMYFFIWIAPRIPTLASAKLDKQVGLWLEGHDEALYQRI 5236  
QY 381 -D-N-C- 383  
Db 5237 MGAELPEDEMGPGAAALTVLRGIWAKVLRNPVEDVKPSKWSLSLGGDSISAMKLLAKCRS 5296  
QY 384 -V-H-D-L-Q- 390  
Db 5297 EGINLNLQILRAKSLSHLAADVKSVIDLHEKEQNDRPALSPIORFYVEAGSIENSTH 5356  
QY 391 -NKE-I-Y-TI-K-W-SP- 401  
Db 5357 FN-QSSTLRILRYQVQAMVORALNSIVECHSMLRARFSDKNNGQWQOOLVMSKVSQSYAFT 5415  
QY 402 -TCP- 404  
Db 5416 AHDVSTASAAGAIISSTOKSLDIRTPGFVAVDLFNLKGHQLFMAAHVLVIDVWSGILL 5475  
QY 405 -G-T-N-N- 408  
Db 5476 GLEDLLGSGPVTTLPRLSPFQWCKMOTSSASEITQUTVKNQPLVVEPANFAPFWMGMDV 5535  
QY 409 -FN-A-N-L-MLASA-SFDS-TV 424  
Db 5536 RPNVYGDSEDFVIDKETSAFNDHNVKTDLDVILIA-AILHSF-SRVFINRKAFTL 5593  
QY 425 -R-LWD-VD-R-G-I-C-IHTL-T-K- 440  
Db 5594 FNESHGREVWGSNLDLSRTVGFYPTVTPIDDEDEVIHTLRQVKDTRKVASNGR 5653  
QY 441 -HQ-E-P-V-Y-SVAFSP-D- 453  
Db 5654 PYFAHRYLTEDGKERFANHAPMEVLFNLYLGRQEQSGHSDSL-LSPTQVEGDDDETSDBGV 5712  
QY 454 -G-RY-LA- 458  
Db 5713 KTSRMALFEISAVTEGQIQLSFMYNRYSKNQKIRRWIABQORTLEEIAELAKTDRPQ 5772  
QY 459 -S-S-SFD-KC-V-HI-W 469  
Db 5773 PTMADPPLPLESYSRLDRVLKTLPHAGVSPFQVEDMYPCSPIDQGMILSQIKSPESY 5832  
QY 470 -NT-QV- 473  
Db 5833 SSTFEVRSKRGPDVASKVVDGWKQVVARHPALRTIFIDSVCKGVFDQIVVKSPPDSGIV 5892  
QY 474 -C-L-H-YLNG-Q-VLLN- 485  
Db 5893 TYKCADELATLLESIRHSSSLGKKKPVLPQOAAVVQTSSTGKIFVKVIVNHAVIDGSLG 5952  
QY 486 -L-GR-S-I-C- 491  
Db 5953 VIGQDLQEAVEGRLSDGPLYSAIKYLRALPAEDAIAWKAKLRGVSPCYFPTTTPRPS 6012  
QY 492 -L-YT- 494  
Db 6013 KPRQLQSLDMRTRFDELHDLAESSNVTTIANILLAAWALVLRSYTNSSDVCYGLTSGRN 6072

QY 495 -----L-----PH-----H----- 498  
Db 6073 VPIDRIENAVGAFINMLVSRIELKSATSLSLEIVENVQSDVFGSMPHQHCSLAQFQHDIGL 6132  
QY 499 -L-VV-I-P-L-V-AL-I- 508  
Db 6133 SGKSLFNTAVSIQRRSSPEELTSDSGIFEQLDGHDPSFAITVNIIDATRDDEGVRFY 6192  
QY 509 -----E-----L-----VL-----K 514  
Db 6193 WSDAVTDGEAKNVSTLMAKILVQVLSNPK 6221

Search completed: January 3, 2005, 15:25:30  
Job time : 131.333 secs

This Page Blank (uspto)





Db 181 SLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240  
QY 241 ATGSYDGFARIWTKDGNLSTLQHKGPFIKWKNNKGNFILSAGVDKTTIIWDAHTGEA 300  
Db 241 ATGSYDGFARIWTKDGNLSTLQHKGPFIKWKNNKGNFILSAGVDKTTIIWDAHTGEA 300  
QY 301 KQPFPHSAPALDQVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGTNEVNAIKWDPT 360  
Db 301 KQPFPHSAPALDQVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGTNEVNAIKWDPT 360  
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPTNNPNANILMASAF 420  
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPTNNPNANILMASAF 420  
QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQTALVHSY 480  
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQTALVHSY 480  
QY 481 RGTGGIFEVCMNAAGDKVGASDGSVCVLDLRK 514  
Db 481 RGTGGIFEVCMNAAGDKVGASDGSVCVLDLRK 514

## RESULT 3

ABO07190  
ID ABO07190 standard; protein; 514 AA.

AC ABO07190;

DT 13-AUG-2003 (first entry)

XX Human p53 modifying protein, SEQ ID 150.

DE Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;  
KW antapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;  
KW lung cancer; ovarian cancer; angiogenesis; cell cycle;  
KW apoptotic disorder; cell proliferation disorder.

OS Homo sapiens.

XX WO200299122-A1.

XX 12-DEC-2002.

XX 03-JUN-2002; 2002WO-US017382.

XX 05-JUN-2001; 2001US-0296076P.

XX 10-OCT-2001; 2001US-0328605P.

XX 15-FEB-2002; 2002US-0357253P.

XX (EXEL-) EXELIXIS INC.

XX Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP;

XX WPI; 2003-156859/15.

XX N-PSDB; ACD13365.

XX Identifying modulators of the p53 pathway for use in treating apoptotic

XX or cell proliferation disorders, comprises screening for agents that

XX modulate activity of a human ortholog of genes that modify the p53

XX pathway in Drosophila.

XX Example 2; Page 458-459; 678pp; English.

XX The invention relates to identifying (M1) a candidate p53 pathway

XX modulating agent, by contacting an assay system comprising a purified HM

XX polypeptide (human ortholog of genes that modify the p53 pathway in

XX Drosophila) or nucleic acid with a test agent under conditions, where but

XX for the presence of the test agent, the system provides a reference

CC sequence, where p53 function is restored), modulating (M3) a p53 pathway  
CC in a mammalian cell (comprising contacting the cell with an agent that  
CC specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4)  
CC a disease in a patient (comprising: (a) obtaining a biological sample  
CC from the patient; (b) contacting the sample with a probe for HM  
CC expression; (c) comparing the results with a control; and (d) determining  
CC whether the comparison indicates a likelihood disease). (M1) is useful  
CC for identifying modulators of the p53 pathway. A probe for HM expression  
CC is useful for diagnosing breast, colon, kidney, lung and ovarian cancer,  
CC in a patient, where the cancer has greater than 25% expression level.  
CC Modulators identified by (M1) are useful in a variety of diagnostic and  
CC therapeutic applications, where disease or disorder prognosis is related  
CC to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell  
CC proliferation disorders (e.g. cancer). Another two new methods (M2 and  
CC M3) are useful for modulating the p53 pathway of a cell, thus restoring  
CC the p53 function of the cell, so that the cell undergoes normal  
CC proliferation or progression through the cell cycle. (M2) and (M3) are  
CC also useful for treating defects in the p53 pathway such as angiogenic,  
CC apoptotic or cell proliferation disorders. The present sequence  
CC represents a human p53 pathway modifying protein  
XX  
SQ Sequence 514 AA;

Query Match 99.5%; Score 3640; DB 6; Length 514;

Best Local Similarity 99.4%; Pred. NO. 1.3e-55;

Matches 511; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MSISDEVNFLVRYLOESGFSHSAFTFGIKSHISOSNINGALVPPAALISIIKGLQYV 60

Db 1 MSISDEVNFLVRYLOESGFSHSAFTFGIKSHISOSNINGALVPPAALISIIKGLQYV 60

QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQOQAAAAAASQ 120

Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVQTRQAYRDKLAQOQAAAAAASQ 120

QY 121 QGSAXNGENTANGEAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVEFICAWN 180

Db 121 QGSAXNGENTANGEAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVEFICAWN 180

QY 181 SLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240

Db 181 SLLASGSGDSTARIWNLSNSTSGTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240

QY 241 ATGSYDGFARIWTKDGNLSTLQHKGPFIKWKNNKGNFILSAGVDKTTIIWDAHTGEA 300

Db 241 ATGSYDGFARIWTKDGNLSTLQHKGPFIKWKNNKGNFILSAGVDKTTIIWDAHTGEA 300

QY 301 KQPFPHSAPALDQVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGTNEVNAIKWDPT 360

Db 301 KQPFPHSAPALDQVQSNNTFASCTDMCIHVCKLGQDRPIKTFQGTNEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPTNNPNANILMASAF 420

Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGPTNNPNANILMASAF 420

QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQTALVHSY 480

Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFPDGRYLASGSPDKCVHIWNTQTALVHSY 480

QY 481 RGTGGIFEVCMNAAGDKVGASDGSVCVLDLRK 514

Db 481 RGTGGIFEVCMNAAGDKVGASDGSVCVLDLRK 514

RESULT 4

ADJ76263

ID ADJ76263 standard; protein; 514 AA.

XX ADJ76263;

XX 20-MAY-2004 (first entry)

XX Marker gene related amino acid sequence SEQ ID NO:1515.

DE

XX	bronchial asthma; chronic obstructive pulmonary disease;
KW	respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;
XX	gene therapy; marker.
XX	
OS	Mus musculus.
XX	
XX	EP1394274-A2.
XX	
XX	03-MAR-2004.
XX	
XX	04-AUG-2003; 2003EP-00254857.
XX	
XX	06-AUG-2002; 2002JP-00299312.
PR	20-MAR-2003; 2003JP-00077212.
XX	
XX	(GENO-) GENOX RES INC.
PA	
XX	
PI	Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;
XX	WPI; 2004-193155/19.
XX	
XX	Testing for bronchial asthma or chronic obstructive pulmonary disease by
PT	comparing the expression level of a marker gene in a biological sample
PT	from a subject with the expression level of the gene in a sample from a
PT	healthy subject.
XX	
XX	Claim 16; SEQ ID NO 1515; 241pp; English.
XX	
XX	The present invention describes a method of testing for bronchial asthma
CC	or chronic obstructive pulmonary disease. The method comprises
CC	determining the expression level of a marker gene in a biological sample
CC	from a subject, comparing the expression level determined with the
CC	expression level of the marker gene in a biological sample from a healthy
CC	subject, and judging whether the subject has bronchial asthma or chronic
CC	obstructive pulmonary disease. The marker gene comprises: (a) a group of
CC	genes (S1) whose expression levels increase when respiratory epithelial
CC	cells are stimulated with interleukin-13; or (b) a group of genes (S2)
CC	whose expression levels decrease when respiratory epithelial cells are
CC	stimulated with interleukin-13. Also described: (1) a reagent (I) for
CC	testing for bronchial asthma or chronic obstructive pulmonary disease;
CC	(2) a kit for screening for a candidate compound for a therapeutic agent
CC	to treat bronchial asthma or chronic obstructive pulmonary disease; (3)
CC	an animal model for bronchial asthma or chronic obstructive pulmonary
CC	disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a
CC	method for producing an animal model for bronchial asthma or chronic
CC	obstructive pulmonary disease; (6) a therapeutic agent for bronchial
CC	asthma or chronic obstructive pulmonary disease, comprising the compound,
CC	a marker gene or an antisense nucleic acid corresponding to a portion of
CC	the marker gene, a ribozyme, a polynucleotide that suppresses the
CC	expression of the gene through an RNAi effect or an antibody recognising
CC	a protein encoded by a marker gene; and (7) a DNA chip for testing for
CC	bronchial asthma or a chronic obstructive pulmonary disease, on which a
CC	probe has been immobilised to assay a marker gene. (I) has respiratory
CC	and antiasthmatic activities, and can be used in gene therapy. The method
CC	is useful for testing for or screening for a therapeutic agent for
CC	bronchial asthma or chronic obstructive pulmonary disease. The present
CC	sequence is used in the exemplification of the present invention.
XX	
XX	Sequence 514 AA;
XX	Sequence 514 AA;

Qy	121	QGS	K	G	E	N	T	A	N	G	E	A	H	T	I	A	N	N	H	D	M	E	V	D	G	V	E	I	P	P	N	K	A	V	L	R	G	H	E	S	E	V	F	I	C	A	N	N	P	V	180											
Db	121	QGS	A	K	G	E	N	T	A	N	G	E	A	H	T	I	A	N	N	H	D	M	E	V	D	G	V	E	I	P	P	N	K	A	V	L	R	G	H	E	S	E	V	F	I	C	A	N	N	P	V	180										
Qy	181	S	D	L	L	A	S	G	S	G	D	T	A	R	I	W	N	L	S	E	N	S	T	G	S	T	O	L	V	L	R	H	C	I	R	E	G	G	D	V	P	S	N	K	D	V	T	S	L	D	N	S	E	G	T	L	240					
Db	181	S	D	L	L	V	S	G	S	G	D	T	A	R	I	W	N	L	S	E	N	S	T	G	S	T	O	L	V	L	R	H	C	I	R	E	G	G	D	V	P	S	N	K	D	V	T	S	L	D	N	S	E	G	T	L	240					
Qy	241	A	T	G	S	Y	D	G	F	A	R	I	W	T	K	D	G	N	L	A	S	T	L	G	H	K	G	P	I	F	A	L	K	W	N	K	K	G	N	F	I	L	S	A	G	V	D	K	T	I	I	W	D	A	H	T	G	E	A	300		
Db	241	A	T	G	S	Y	D	G	F	A	R	I	W	T	K	D	G	N	L	A	S	T	L	G	H	K	G	P	I	F	A	L	K	W	N	K	K	G	N	F	I	L	S	A	G	V	D	K	T	I	I	W	D	A	H	T	G	E	A	300		
Qy	301	K	Q	P	P	F	H	S	A	P	A	L	D	V	D	W	S	N	N	T	F	A	S	C	S	T	D	M	C	I	H	V	K	L	G	D	R	I	K	T	F	Q	G	H	T	N	E	V	N	A	I	K	W	D	P	T	360					
Db	301	K	Q	P	P	F	H	S	A	P	A	L	D	V	D	W	S	N	N	T	F	A	S	C	S	T	D	M	C	I	H	V	K	L	G	D	R	I	K	T	F	Q	G	H	T	N	E	V	N	A	I	K	W	D	P	T	360					
Qy	361	G	N	L	A	S	C	S	D	D	T	L	K	I	W	S	M	K	D	N	C	V	H	D	L	Q	A	H	N	K	E	I	Y	T	I	K	W	S	T	G	P	G	T	N	N	P	N	A	M	L	A	S	A	S	F	420						
Db	361	G	N	L	A	S	C	S	D	D	T	L	K	I	W	S	M	K	D	N	C	V	H	D	L	Q	A	H	N	K	E	I	Y	T	I	K	W	S	T	G	P	G	T	N	N	P	N	A	M	L	A	S	A	S	F	420						
Qy	421	D	S	T	V	R	L	M	D	V	D	R	G	I	C	I	H	T	L	T	K	H	Q	E	P	V	Y	S	V	A	F	S	P	D	G	R	Y	L	A	S	G	S	F	D	K	C	V	H	I	W	N	T	O	T	G	A	L	V	H	S	Y	480
Db	421	D	S	T	V	R	L	M	D	V	D	R	G	I	C	I	H	T	L	T	K	H	Q	E	P	V	Y	S	V	A	F	S	P	D	G	R	Y	L	A	S	G	S	F	D	K	C	V	H	I	W	N	T	O	T	G	A	L	V	H	S	Y	480
Qy	481	R	G	T	G	G	I	F	E	V	C	W	N	A	A	G	D	K	V	G	A	S	A	S	D	G	S	V	C	V	L	D	R	K	514																											
Db	481	R	G	T	G	G	I	F	E	V	C	W	N	A	A	G	D	K	V	G	A	S	A	S	D	G	S	V	C	V	L	D	R	K	514																											

RESULT 5  
ADD14051

ADD14051  
ID ADD14051 standard; protein; 577 AA.

AA  
AC ADD14051;

01-JAN-2004 (first entry)

DE Human src biomarker polypeptide SEQ ID NO:240.

XX predictor set; protein tyrosine kinase activity modulator;  
KW protein tyrosine kinase pathway; protein tyrosine kinase; cytosstatic;  
KW gene therapy; drug sensitivity; genetic profile; cancer; human.

XX Homo sapiens.

XX  
PN  
WO2003062395-A2

31-JUL-2003

XX  
PF 17-JAN-2003: 2003WO-US001981.

18-JAN-2002 2002US-0350061P

XX  
BA (BRIM ) BRISTOL-MYERS SQUIBB CO

XX  
DT Huang R Fairchild CP Lee FY Shaw P.

XX  
DB WPT: 2003-636735/60

BR WP1; 2003-636/35/  
DR N-PSDB: ADD14646.

XX	PT	PT	PT
			New polynucleotides and polypeptides for predicting the activity of compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways.

XX  
PS Claim 10: SEO ID NO 240: 139pp: English.

xx The present invention describes a predictor set comprising a plurality of  
cc polypeptides or polypeptides whose expression pattern is predictive of  
cc the response of cells to treatment with a compound that modulates protein  
cc tyrosine kinase activity or members of the protein tyrosine kinase  
cc pathway. Also described: (1) predicting whether a compound is capable of  
cc modulating the activity of cells, comprising obtaining a sample of cells,  
cc determining whether the cells express a plurality of markers, and



correlating the expression of the markers to the compound's ability to modulate the activity of the cells; (2) a plurality of cell lines for identifying polynucleotides and polypeptides whose expression levels correlate with compound sensitivity or resistance of cells associated with a disease state; and (3) identifying polynucleotides and polypeptides that predict compound sensitivity or resistance of cells associated with a disease state by using the expression pattern of the microarray. The polynucleotides and polypeptides have cytosolic activities, and can be used in gene therapy. The polynucleotides and polypeptides are useful in predicting the activity of compounds that interact with protein tyrosine kinases and/or protein tyrosine kinase pathways. These may be used in determining drug sensitivity in patients to allow the development of individualized genetic profiles which aid in treating diseases and disorders (e.g. cancer) based on patient response at a molecular level. The present sequence is used in the exemplification of the present invention.

XX Sequence 577 AA;

Query Match 91.7%; Score 3355.2; DB 7; Length 577;  
Best Local Similarity 86.2%; Pred. No. 6.2e-50;  
Matches 456; Conservative 38; Mismatches 17; Indels 18; Gaps 10;

QY 1 MSISDEVNVLVRYLQESGFSAFTFGIKSHISQSNINGALVPPAALISITQKGLQYV 60  
DB 52 MSITSDEVNVLVRYLQESGFSAFTFGIESHSQSNINGTLVPPAALISITQKGLQYV 111  
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAVRDKLAQQ--AAAAAAA-AA 117  
DB 112 EAEISINEDGTVDPGRPIESLSLIDAVMPDVVQTRQAFREKLAQQASAAAAAATAA 171  
QY 118 ASQ-----QG-S-----AKGENTANGEANGAHTIANNHTDM--MEVDGDVEIPPNKAVLVR 166  
DB 172 ATAATTTSAGVSHQNPKNREATVNGEENRAHSV--NNHA--KPEIDGVEIIPSKATVLR 229  
QY 167 GHESEVFIQAWNPVSDLLASGSDSTARINWLNSENSTGSLVLRHCIREGGQDVPNSK 226  
DB 230 GHESEVFIQAWNPVSDLLASGSDSTARINWLNSENNGSGSTQVLRHCIREGGHDPVPSN 289  
QY 227 DVTSLDWNSBETLLATGSDYDGFARITWKGNLSTLQHKGPFPALKNNKGNFILLSAGV 286  
DB 290 DVTSLDWNTNGTLLATGSDYDGFARITWEDGNLSTLQHKGPFPALKNNKGNFILLSAGV 349  
QY 287 DKTTIIWDAHTGEAKQOFPFHSAPALDQVDSNN--TFASCSTDMCHVCKLGDPRPKTF 345  
DB 350 DKTTIIWDAHTGEAKQOFPFHSAPALDQVDSNN--NTTTFASCSTDMCHVCKLGDPRPKTF 408  
QY 346 QGHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGP 405  
DB 409 QGHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 468  
QY 406 TNNPNANMLASFSSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSGFDKC 465  
DB 469 TSNPNINMLASFSSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLAGSGFDKC 528  
QY 466 VHIWNTQTGLVHYSYRGTTGGIPVPCWNAAGDKVGASDGSVCVLDLRK 514  
DB 529 VHIWNTQSGNLVHYSYRGTTGGIPVPCWNAAGDKVGASDGSVCVLDLRK 577

RESULT 6  
ADQ18019  
ID ADQ18019 standard; protein; 577 AA.

XX ADQ18019;

AC ADQ18019;

XX 26-AUG-2004 (first entry)

XX Human soft tissue sarcoma-upregulated protein - SEQ ID 836.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
XX Homo sapiens.  
XX WO2004048938-A2.  
XX 10-JUN-2004.  
XX 26-NOV-2003; 2003WO-US038193.  
XX 26-NOV-2002; 2002US-0429739P.  
XX (PROT-) PROTEIN DESIGN LABS INC.  
XX Aziz N, Ginsburg WM, Zlotnik A;  
XX WPI; 2004-441208/41.  
XX Early detection of soft tissue sarcoma comprises determining expression of a gene in a first soft tissue sample and a normal soft tissue sample and comparing the gene expression, also useful in treating soft tissue sarcoma.

Example 2; SEQ ID NO 836; 210pp; English.

The invention relates to a novel method for detecting soft tissue sarcoma which comprises obtaining a first soft tissue sample from an individual and a normal soft tissue sample from the same or different individual, determining the expression of a gene in both samples and comparing the expression of the gene in both soft tissue samples, where a higher level of protein expression in the first soft tissue sample indicates the presence of soft tissue sarcoma. The method of the invention has cytosolic applications and may be useful for detecting soft tissue sarcoma, possibly via gene therapy or vaccine production. The nucleic acid sequences may be useful in diagnostic and screening applications. The current sequence is that of a human soft tissue sarcoma-upregulated protein of the invention. The current sequence is not shown within the specification per se but was submitted in CD format by the inventor.

Sequence 577 AA;

Query Match 91.7%; Score 3355.2; DB 8; Length 577;  
Best Local Similarity 86.2%; Pred. No. 6.2e-50;  
Matches 456; Conservative 38; Mismatches 17; Indels 18; Gaps 10;

QY 1 MSISDEVNVLVRYLQESGFSAFTFGIKSHISQSNINGALVPPAALISITQKGLQYV 60  
DB 52 MSITSDEVNVLVRYLQESGFSAFTFGIESHSQSNINGTLVPPAALISITQKGLQYV 111  
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAVRDKLAQQ--AAAAAAA-AA 117  
DB 112 EAEISINEDGTVDPGRPIESLSLIDAVMPDVVQTRQAFREKLAQQASAAAAAATAA 171  
QY 118 ASQ-----QG-S-----AKGENTANGEANGAHTIANNHTDM--MEVDGDVEIPPNKAVLVR 166  
DB 172 ATAATTTSAGVSHQNPKNREATVNGEENRAHSV--NNHA--KPEIDGVEIIPSKATVLR 229  
QY 167 GHESEVFIQAWNPVSDLLASGSDSTARINWLNSENSTGSLVLRHCIREGGQDVPNSK 226  
DB 230 GHESEVFIQAWNPVSDLLASGSDSTARINWLNSENNGSGSTQVLRHCIREGGHDPVPSN 289  
QY 227 DVTSLDWNSBETLLATGSDYDGFARITWKGNLSTLQHKGPFPALKNNKGNFILLSAGV 286  
DB 290 DVTSLDWNTNGTLLATGSDYDGFARITWEDGNLSTLQHKGPFPALKNNKGNFILLSAGV 349  
QY 287 DKTTIIWDAHTGEAKQOFPFHSAPALDQVDSNN--TFASCSTDMCHVCKLGDPRPKTF 345  
DB 350 DKTTIIWDAHTGEAKQOFPFHSAPALDQVDSNN--NTTTFASCSTDMCHVCKLGDPRPKTF 408  
QY 346 QGHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQDNCVHDLOQHNKEIYTIKWSPTGP 405  
DB 409 QGHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 468



PR 17-NOV-2000; 2000US-0249211P.  
 PR 17-NOV-2000; 2000US-0249212P.  
 PR 17-NOV-2000; 2000US-0249213P.  
 PR 17-NOV-2000; 2000US-0249214P.  
 PR 17-NOV-2000; 2000US-0249215P.  
 PR 17-NOV-2000; 2000US-0249216P.  
 PR 17-NOV-2000; 2000US-0249217P.  
 PR 17-NOV-2000; 2000US-0249218P.  
 PR 17-NOV-2000; 2000US-0249244P.  
 PR 17-NOV-2000; 2000US-0249245P.  
 PR 17-NOV-2000; 2000US-0249264P.  
 PR 17-NOV-2000; 2000US-0249265P.  
 PR 17-NOV-2000; 2000US-0249297P.  
 PR 17-NOV-2000; 2000US-0249299P.  
 PR 01-DEC-2000; 2000US-0250160P.  
 PR 01-DEC-2000; 2000US-0250391P.  
 PR 05-DEC-2000; 2000US-0251030P.  
 PR 05-DEC-2000; 2000US-0251988P.  
 PR 06-DEC-2000; 2000US-0256719P.  
 PR 06-DEC-2000; 2000US-0251479P.  
 PR 08-DEC-2000; 2000US-0251856P.  
 PR 08-DEC-2000; 2000US-0251868P.  
 PR 08-DEC-2000; 2000US-0251869P.  
 PR 08-DEC-2000; 2000US-0251989P.  
 PR 11-DEC-2000; 2000US-0251990P.  
 PR 05-JAN-2001; 2001US-0259678P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 XX Rosen CA, Barash SC, Ruben SM;  
 XX  
 XX WPI; 2001-476159/51.  
 DR N-PSDB; ADML19381.  
 XX  
 PT Isolated nucleic acid molecule encoding a channel/transporter protein is  
 used in preventing, treating or ameliorating a medical condition.  
 XX  
 PS Claim 11; SEQ ID NO 667; 809pp; English.  
 XX  
 CC The invention relates to an isolated nucleic acid molecule encoding a  
 channel/transporter protein or sequences at least 95% identical to a  
 these. The nucleic acids and proteins encoded by them are used to  
 prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
 rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
 in diagnosing a pathological condition or susceptibility to a  
 pathological condition. The antibodies to the proteins can also be used  
 in alleviating symptoms associated with the disorders and in diagnostic  
 immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
 (ELISA). Disorders which are diagnosed or treated include autoimmune  
 diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
 neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
 arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,  
 nervous system disorders e.g. Alzheimer's disease, infections caused by  
 bacteria, viruses and fungi and ocular disorders e.g. corneal infection.  
 The polypeptides can also be used to aid wound healing and epithelial  
 cell proliferation, to prevent skin aging due to sunburn, to maintain  
 organs before transplantation, for supporting cell culture of primary  
 tissues, to regenerate tissues and in chemotaxis. The polypeptides can  
 also be used as a food additive or preservative to increase or decrease  
 storage capabilities. This sequence corresponds to a protein of the  
 invention.  
 XX  
 SQ Sequence 459 AA;  
 Query Match 89.7%; Score 3282.7; DB 4; Length 459;  
 Best Local Similarity 92.9%; Pred. No. 5e-49;  
 Matches 457; Conservative 2; Mismatches 0; Indels 33; Gaps 3;  
 23 HSAFTFGIKSHISQSNINGALVPPAALISIQKLOYVEAEVSNEDGTLFDCGRPIESLS 82  
 1 H-A-----SD-----VEAEVSNEDGTLFDCGRPIESLS 27

QY 83 LIDAVNPVVQTRQAYRDKLAQQQAAAAAAAAAAAAAQQGSAKNGENTANGENGAAHTIA 142  
 DB 28 LIDAVNPVVQTRQAYRDKLAQQQAAAAAAAAAAAAAQQGSAKNGENTANGENGAAHTIA 87  
 QY 143 NNHTDMMEVDGVEIPPNKAVVLRGHESEVFI CAMNPVSDLLASGSGDSTARINWLSNS 202  
 DB 88 NNHTDMMEVDGVEIPPNKAVVLRGHESEVFI CAMNPVSDLLASGSGDSTARINWLSNS 147  
 QY 203 TSGSTQLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLLATGSDYDGFARIWTKDGNLSTL 262  
 DB 148 TSGSTQLVLRHCIREGGQDVPSNKNVTSLDWNSSEGLLATGSDYDGFARIWTKDGNLSTL 207  
 QY 263 GQHKGFIFALKWNNKGNFILLSAGVDKTTIWDHTGEAKQOFPFHSAPALDWDQSNNTF 322  
 DB 208 GQHKGFIFALKWNNKGNFILLSAGVDKTTIWDHTGEAKQOFPFHSAPALDWDQSNNTF 267  
 QY 323 ASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGTNLLASCSDMTLKIWSMKQDN 382  
 DB 268 ASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPGTNLLASCSDMTLKIWSMKQDN 327  
 QY 383 CVHDLQHNKEIYTIKWSPTGCTNNPNANMLASASFDSTVRLWDVDRGICHTLTQH 442  
 DB 328 CVHDLQHNKEIYTIKWSPTGCTNNPNANMLASASFDSTVRLWDVDRGICHTLTQH 387  
 QY 443 EPVYSVAFSPDGRYLASGSGFDKCVHIWNTQTGALVHSYRGTGGIFEVCAAGDKVGASA 502  
 DB 388 EPVYSVAFSPDGRYLASGSGFDKCVHIWNTQTGALVHSYRGTGGIFEVCAAGDKVGASA 447  
 QY 503 SDGSVCVLDLRK 514  
 DB 448 SDGSVCVLDLRK 459  
 RESULT 8  
 ABP41760  
 ID ABP41760 standard; protein; 542 AA.  
 XX  
 AC ABP41760;  
 XX  
 DT 22-AUG-2002 (first entry)  
 XX Human ovarian antigen HNOKM38, SEQ ID NO:2892.  
 DE  
 DE Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive.  
 XX Homo sapiens.  
 OS  
 XX WO200200677-A1.  
 PN  
 XX 03-JAN-2002.  
 PD  
 PD 07-JUN-2001; 2001WO-US018569.  
 PF  
 XX 07-JUN-2000; 2000US-0209467P.  
 PR (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PA Birse CE, Rosen CA;  
 XX WPI; 2002-147878/19.  
 DR N-PSDB; ABQ54837.  
 XX  
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,

useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian cancer), immune disorders, cardiovascular disorders and neurological diseases.

Claim 11; SEQ ID NO 2892; 2922pp; English.

The invention relates to 2175 novel human ovarian antigens (ABP41054-ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also encompasses polypeptides 90% identical and polynucleotides 95% identical to the sequences of the invention. The invention additionally relates to recombinant vectors and host cells comprising human ovarian antigen polynucleotides, antibodies against human ovarian antigens, and the use of ovarian antigen polynucleotides and polypeptides in diagnosing, treating, prognosing or preventing various ovary and/or breast-related disorders. Such conditions include ovarian cancer and breast cancer, and metastatic tumours of ovarian or breast origin, reproductive system disorders (e.g., infertility, disorders of pregnancy, anovulation, polycystic ovary syndrome, ovarian cysts, and dysmenorrhoea), endocrine disorders, infections (e.g., chlamydia, HIV, toxoplasmosis, and toxic shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and vaginitis), immune disorders (e.g., congenital and acquired immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus), blood-related disorders (e.g., anaemia), cardiovascular disorders, respiratory disorders, neurological disorders, gastrointestinal disorders and urinary system disorders. Ovarian antigen polypeptides and polynucleotides may also be used in screening for compounds which modulate ovarian antigen expression or activity. The polynucleotides may further be used for gene therapy, chromosome mapping, in the identification of individuals and in forensic analysis, and the polypeptides may be used as food additives or to prepare antibodies useful in disease diagnosis, drug targeting and phenotyping. The present sequence represents a human ovarian antigen of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 542 AA;

Query Match 89.2%; Score 3265.6; DB 5; Length 542;  
Best Local Similarity 80.7%; Pred. No. 2.4e-48;  
Matches 442; Conservative 42; Mismatches 20; Indels 44; Gaps 15;

```
QY 5 SDEWVF---LV-----YR-----VLQESGFSAFTFGIKSHISQSNNG 41
DB 1 TDC--FKSKVDLQSQKPKRTPHTLCGCTGFLTC-GFSHAFYFXIESHISQSNNG 57
QY 42 ALVPPAALISIIKGLQYVBAEVSINEDGTLFQGRPIESLSLIDAVMPDVVQTRQAYRD 101
DB 58 TLVPPAALISIIKGLQYVBAEVSINEDGTVFDGRPIESLSLIDAVMPDVVQTRQAFRE 117
QY 102 KLAQOO--AAAAA--AAASQ-----QG-S-----AKNGENTANGEAGHTIANHTDM 148
DB 118 KLAQOOASAAAAAATAATAATTTAGVSHQNPQKREATVNGENRAHSV--NNHA-K 175
QY 149 -MEVDGVEIIPPKAVLVRGHESEVFTICANNPVSDDLASGSGDSTARINWLSNENSGST 207
DB 176 PMIDGEVEIPSSKATVLRGHESEVFTICANNPVSDDLASGSGDSTARINWLSNENSGST 235
QY 208 QLVLRHCIREGGQDPSPKNDVTLSDMNSGFTLLATGSDYGFARIWTKDGNLSTLQGHKG 267
DB 236 QLVLRHCIREGGHVPSPKNDVTLSDMNTGNTLLATGSDYGFARIWTEGDLNASTLQGHKG 295
QY 268 PIFALKWKNKGNFTLSAGVDKTLIIWDAHTGEAKQPPHSPALDQVDSQNN--TFASCS 326
DB 296 PIFALKWNRKGNFTLSAGVDKTLIIWDAHTGEAKQPPHSPALDQVDSQNN--NNTTFASCS 354
QY 327 TDMCIHVCKLQGRPIKTFQGHTEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNCVHD 386
DB 355 TDMCIHVCRIGCBPVTFTQGHTEVNAIKWDPSGMLLASCDDMTLKIWSMKQEVCIHD 414
QY 387 LQOHKNEIYTKWSPGPTGNPNANIMLASGSDSTVRLWDVDRGICHTTLTKHQBPVY 446
DB 415 LQAHKNEIYTKWSPGPTGNPNINMLASGSDSTVRLWDIERGVCTHTLTKHQBPVY 474
```

```
QY 447 SVAFSPDGRYLASGSPDKCVHIWNTOTGALVHSYRGTGGIFVCMNAAGKVCASASDGS 506
DB 475 SVAFSPDGRYLASGSPDKCVHIWNTOTGALVHSYRGTGGIFVCMNAAGKVCASASDGS 534
QY 507 VCVLDLRLK 514
DB 535 VCVLDLRLK 542
```

RESULT 9  
ABB60376

ID ABB60376 standard; protein; 700 AA.

XX ABB60376;

AC ABB60376;

DT 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 7920.

DE Drosophila; developmental biology; cell signalling; insecticide;

KW pharmaceutical.

KW Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US009231.

XX 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX (PEKE ) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

PI WPI; 2001-656860/75.

DR N-PSDB; ABL04479.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions.

PT Disclosure; SEQ ID NO 7920; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA sequences (ABU16176-ABU16175) and the encoded proteins (ABU57737-ABU72072). The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 700 AA;

Query Match 79.7%; Score 2917.1; DB 4; Length 700;  
Best Local Similarity 57.4%; Pred. No. 3.7e-41;  
Matches 411; Conservative 50; Mismatches 36; Indels 219; Gaps 58;

```
QY 1 MSTSSEVNFPLVTRYLOESGFSAFTFGIKSHISQSNNGALVPPAALISIIKGLQYV 60
DB 1 MSTSSEVNFPLVTRYLOESGFSAFTFGIKSHISQSNNGALVPPAALISIIKGLQYV 60
```

```
QY 61 EAEVSINEDGTLFD-GRPIESLSLIDAVMPDV-----VOTRQ-----KLA-- 104
```

```
DB 61 EAEVSINEDGTLFD-GRPIESLSLIDAVMPDV-----VOTRQ-----KLA-- 114
```

```
QY 105 -Q-----Q-----Q-A-----AA-----A 112
```

Db 115 NQNNNAKPEIKBPGTGVAGSGCKIAGSTTCTGTPTDQSAASEVDSGNAANAGGTVA 174  
QY 113 ----A-----A-----AA-ASQ-----Q-----CSAK--N-G-EN--T-- 130  
Db 175 GNNAGGNOASTGCGNSTSTPAGDGLAAGASOKKQNSNEAGSSGSGNAGNANATSTD 234  
QY 131 -A-----NG-----EE-----N-----G-AH-----TI- 141  
Db 235 AASSTNGNSSSSVEQPTSLGTAGGTVTSNPDAAASGASATGSKAPSGAVTIR 294  
QY 142 --A--NN-----H-----T-DMEV--DGD--VRIPPNKAIVLRG 167  
Db 295 VCAQGNVQSGSSNAQSSAPSGTISSTSGGAGTAAAL-VPMDIDENIBIPESKARVLRG 353  
QY 168 HSEVEPICAWNPVS-DLLASGSDSTARINLSE-NSTSGSTOLVLRHCIREGQDVPSN 225  
Db 354 HSEVEFICAWNP-SRDLASGSDSTARIDMSDAN-T-NSNQLVLRHCIRGGAGVPSN 410  
QY 226 KDVTSLDMNSEGTLATGSDGPARIW-TKDGNLASTLGOHKGPFPALKWKK-GNFILS 283  
Db 411 KDVTSLDMNDCGSLATGSDGYARIWKT-DGRLASTLGOHKGPFPALKWN-KCGNYILS 468  
QY 284 AGVDKTTIWDATGE-AKQPPFHPAPALVDVQSNNTFASGSTD-MCHVCKLG--QD 339  
Db 469 AGVDKTTIWDASTGQCT-QQAFHAPALVDVQTNQAFASGSTDQR-IHVCRLGWNE- 525  
QY 340 RPIKTFQHTNEVNAIKWDPNTGNLLASCSDDMTLKIWSMKQD-NCVHDLQHNKEIYTIK 398  
Db 526 -PIKTFQHTNEVNAIKWCPQGLLASCSDDMTLKIWSMNRDCC-HDLQAHKSEIYTIK 583  
QY 399 WSPGTGTTNNPNANMLASASFDSTVRLVDVDRGICHTLTKHOEVPVSVAFSPDGRYLA 458  
Db 584 WSPGTGTTNNPNANMLASASFDSTVRLVDVDRGICHTLTKHOEVPVSVAFSPDGRYLA 458  
QY 459 SSGFDCVCHVNTQTGALVHSYRGTCGIFEVCMNAAGDKVGASASDGSVCVLDLRK 514  
Db 644 SSGFDCVCHVNTQTGALVHSYRGTCGIFEVCMNSKGTGVASASDGSVCVLDLRK 699

RESULT 10  
ABG21351  
ID ABG21351 standard; protein; 584 AA.  
XX AC ABG21351;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #21342.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US0008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX PA (HYSE-) HYSEQ INC.  
XX PA Drmanac RT, Liu C, Tang YT;  
XX PI WPI: 2001-639362/73.  
XX DR N-PSDB; AAS85538.  
XX PT New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, gene mapping, identification of mutations.

PT responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
PS Claim 20; SEQ ID NO 51710; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
sequences. (I) is useful as hybridisation probes, polymerase chain  
reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological actions in  
CC polypeptide and polynucleotide sequences have application of mutations  
CC diagnostics, for gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences of the invention. Note: The sequence data for this  
CC patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 584 AA;  
Query Match 77.0%; Score 2816.1; DB 4; Length 584;  
Best Local Similarity 63.8%; Pred. No. 1.3e-39;  
Matches 387; Conservative 44; Mismatches 37; Indels 139; Gaps 54;  
QY 2 SI-----S-----D-----EV---N---FLV-YRY-LOE-SGFSSHATFGIKSH 33  
Db 22 AIGAVGRSOASTLAGDPDNGVCVHAGNSRIF-VPLPVGL-EVKGFSHATFGIESH 79  
QY 34 ISQSNINGALVPPAALISIIQKGLQYVEAEVSNEDGTLFDGRPIESLSLIDVAMPDVQ 93  
Db 80 ISQSNINGTLVPPSALISILQKGLQYVEAEISINKDGTVPDSRPIESLSLIVAVPDVQ 139  
QY 94 TRQAYRDKLAQQA-AAA--A-A-AA-----AA-SQGSANKNGENTANGENGAAHTIA 142  
Db 140 MRQQAQGEKLTQQAASAAATEASAKAATMTAPALISQNPKNREATVNGEANGAEI- 198  
QY 143 NNHTDM-MEVDGDVEIPPNKAVLRGHESEVFCANVPVSDLLASGSGDSTARINLSEN 201  
Db 199 NNHS-KPMEIDGDVEIPPNKATVLRGHESEVFCANVPVSDLLASGSGDSTARINLSEN 257  
QY 202 STSGSTQLVLRHCIREGQDVPSNKDVTSLDWNSEGTLLATGSDYDGFARITWKDGNLAST 261  
Db 258 SNGSGTQLVLRHCIREGHDVPSNKDVTSLDWNSDGTLLAMGSDYDGFARITWE--N-A-- 312  
QY 262 LGOHKGPFPAL--KW-NKKGN--FILS-AGVDKTTIWDAT--GEAKQOQFFHSAAPAL 312  
Db 313 -----P--ALDVDMQN---NMTFA-SCS-TD---CI---HVCRLGCD-----H--P-- 344  
QY 313 DVD-WQSNNTFASCSDDMTLKIWSMKQD-NCVHDLQHNKEIYTIK 347  
Db 345 -VKTFO-GHTF--C-T--CIESICFWGGLRKLKTTMTGKRLRP-KTFCSDGGALLPPAGR 396  
QY 348 --H--T-----NEVNAIKWDPNTGNLLASCSDDMTLKIWSMKQD-NCVHDLQ 389  
Db 397 RPHLLTGPDIFPKRLSALFQNEVNAIKWDPNSGMLLASCSDDMTLKIWSMKQDACHDLQA 456  
QY 390 HNKEIYTIKWSPTGPTNNPNANMLASASFDSTVRLVDVDRGICHTLTKHOEVPVSV 449  
Db 457 HSKEIYTIKWSPTGPTNNPNANMLASASFDSTVRLVDVDRGICHTLTKHOEVPVSV 516  
QY 450 FSPDGRYLASGSDKCVHIVNTQTGALVHSYRGTCGIFEVCMNAAGDKVGASASDGSVCV 509  
Db 517 FSPDGRYLASGSDKCVHIVNTQTGALVHSYRGTCGIFEVCMNAAGDKVGASASDGSVCV 576

Qy 510 L-DLR-K 514  
 Db 577 LWQLSGK 583

RESULT 11  
 ABP51424  
 ID ABP51424 standard; protein; 395 AA.  
 XX  
 AC ABP51424;  
 XX  
 XX  
 DT 03-SEP-2002 (first entry)  
 XX  
 DE Human MDDT SEQ ID NO 446.  
 XX  
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;  
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;  
 KW hepatotropic; antiinflammatory; antipsoriatic; cytostatic; anti-HIV;  
 KW antiallergic; antianaemic; antiaethmatic; antiatherosclerotic; antiout;  
 KW neuroprotective; antitirneumatic; antiarthritic.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200240715-A2.  
 XX  
 PD 23-MAY-2002.  
 XX  
 XX 06-SEP-2001; 2001WO-US027628.  
 XX  
 PR 05-SEP-2000; 2000US-0229747P.  
 PR 05-SEP-2000; 2000US-0229748P.  
 PR 05-SEP-2000; 2000US-0229749P.  
 PR 05-SEP-2000; 2000US-0229750P.  
 PR 05-SEP-2000; 2000US-0229751P.  
 PR 05-SEP-2000; 2000US-0230583P.  
 PR 06-SEP-2000; 2000US-0230505P.  
 PR 06-SEP-2000; 2000US-0230514P.  
 PR 06-SEP-2000; 2000US-0230515P.  
 PR 06-SEP-2000; 2000US-0230518P.  
 PR 06-SEP-2000; 2000US-0230519P.  
 PR 06-SEP-2000; 2000US-0230595P.  
 PR 06-SEP-2000; 2000US-0230597P.  
 PR 06-SEP-2000; 2000US-0230598P.  
 PR 06-SEP-2000; 2000US-0230599P.  
 PR 06-SEP-2000; 2000US-0230610P.  
 PR 06-SEP-2000; 2000US-0230610P.  
 PR 06-SEP-2000; 2000US-0230865P.  
 PR 06-SEP-2000; 2000US-0230988P.  
 PR 06-SEP-2000; 2000US-0230989P.  
 PR 07-SEP-2000; 2000US-0230951P.  
 PR 07-SEP-2000; 2000US-0231163P.  
 PR 07-SEP-2000; 2000US-0231167P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;  
 PI Hallman JL, Jones AL, Yu JY, Wright RJ, Gietzen D, Liu TF, Yap PE;  
 PI Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;  
 PI Roseberry AM, Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V;  
 PI Daffo A, Marwaha R, Chen AJ, Chang SC, Au AP, Inman RR;  
 XX  
 DR WPI; 2002-527544/56.  
 DR N-PSDB; ABQ72641.  
 XX  
 XX Novel human disease detection and treatment polypeptide, useful in  
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.  
 PT AIDS.  
 XX  
 XX Claim 14; Page 578; 618pp; English.  
 PS  
 XX

CC The invention relates to an isolated human disease detection and  
 CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a  
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
 CC specification, a naturally occurring polypeptide comprising a sequence  
 CC having at least 90% identity to (I) or a biologically active or  
 CC immunogenic fragment of (I). (I) is useful for screening a compound for  
 CC effectiveness as an agonist or antagonist, for screening a compound that  
 CC specifically binds (I) or modulates the activity of (I), and for  
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
 CC screening a compound for effectiveness in altering expression of a target  
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,  
 CC in a diagnostic test for a condition or a disease associated with the  
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,  
 CC and for purifying (I) from a sample. A composition comprising (I), an  
 CC agonist or antagonist is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional MDDT. (I)  
 CC or (II) are useful for diagnosing, treating or preventing disorders  
 CC associated with aberrant expression of MDDT, where the disorders are  
 CC selected from a cell proliferative disorder such as arteriosclerosis,  
 CC cirrhosis, hepatitis, psoriasis, and cancer and an  
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases, in somatic or  
 CC germline gene therapy, to generate a transcript image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences  
 XX  
 SQ Sequence 395 AA;

Query Match 75.0%; Score 2744; DB 5; Length 395;  
 Best Local Similarity 99.5%; Pred. No. 4.7e-39;  
 Matches 389; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MSISDEVNPLVYRYLOESGFSHTFGIKSHISQSNINGALVPPAALISITKGLQYV 60  
 Db 5 MSISDEVNPLVYRYLOESGFSHTFGIKSHISQSNINGALVPPAALISITKGLQYV 64  
 Qy 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVOTROQAVRDKLAQOAAAAAASQ 120  
 Db 65 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVOTROQAVRDKLAQOAAAAAASQ 124  
 Qy 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFCANPV 180  
 Db 125 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVVLRGHESEVFCANPV 184  
 Qy 181 SLLASGSGDSTARINWLSNSTSGSTQVLRHCIREGGQDVPSNKDVTSLDWNSEGLL 240  
 Db 185 SLLASGSGDSTARINWLSNSTSGSTQVLRHCIREGGQDVPSNKDVTSLDWNSEGLL 244  
 Qy 241 ATGSDGFARITWKGNLSTLQHKGPFPALKWKNKGNFILSAGVDKTTIWDATGEA 300  
 Db 245 ATGSDGFARITWKGNLSTLQHKGPFPALKWKNKGNFILSAGVDKTTIWDATGEA 304  
 Qy 301 KQGFPHSPALPDVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360  
 Db 305 KQGFPHSPALPDVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 364  
 Qy 361 GNLLASCSDDMTLKISWMKQDNCVHDLQOHN 391  
 Db 365 GNLLASCSDDMTLKISWMKQDNCVHDLQOHN 395

RESULT 12  
 ABP60991  
 ID ABP60991 standard; protein; 5635 AA.  
 XX  
 AC ABP60991;  
 XX



QY 104 ---A---Q--QO-----AA----- 109  
Db 961 VASVAGTNNKTSVVHVLPFIHQGOILSTIEGIPVTLPCASGNPKPSVIMSKGEL 1020  
QY 110 ---AAA--A-A-----A-AA--A----- 118  
Db 1021 ISTSAKFSAGADGSLVYVSPGSEGEYVCTATNTAGYAKRKVOLTYYVPRVFGDORG 1080  
QY 119 -SQ-----Q-----GS----- 123  
Db 1081 LSQDKPVEISVLAGEEVLPCFVKSLPPIITWAKETQLISPFSPRHTFLPSCMKITET 1140  
QY 124 ---A---K-----N----- 126  
Db 1141 RTSDSGMYLCVATNIAGNVQAVKLVHVHPKIQRGPKHLKVQGVQVDPICNAQGTPLP 1200  
QY 127 ---GE--N--T---A-----N-G----- 133  
Db 1201 VIITWSKGGTMLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDEITLHVQE 1260  
QY 134 ---E---E--N-----GA-----H----- 139  
Db 1261 PPTVEDLEPPYNTTFOERVANQRIEFPCKPAGTPKPTIKLHNGRELTGREPGISILEDG 1320  
QY 140 T---IA-----N--N--HT---DM----- 148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHPPVVKQEQVNVSVLLNQLTNL 1380  
QY 149 ---M-----E----- 150  
Db 1381 FCEVEGTPSPLIMYKDNQVQTESSTIQTVNNGKILKLPATPEDAGRYSCAKAINAGTS 1440  
QY 151 ---VD---G-----DV--E-----I-----P-----PN 160  
Db 1441 QKYFNIDVLVPTIIGTFNFEVSVLNRDVALEQVKGTFPDIHWFKDGKFLGDPN 1500  
QY 161 ---K-A-----V-----VL 165  
Db 1501 VELLDRGOVLHLKNARNDKRGVQCTVSNAGKQAKDKIKLITYNPSPKGNVTTDISVL 1560  
QY 166 ---RG-----H----- 168  
Db 1561 INSLIKLECETRGLPMPAITWYKDGQPIMSSSQALYIDKQYLHI PRAQVSDSATYCHV 1620  
QY 169 ---E-S-----E-----C--A-----W----- 177  
Db 1621 ANVAGTAKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECKAAGNPSPIITWLKGV 1680  
QY 178 ---N-----P-V--SD-- 182  
Db 1681 PVKANDNIRIEAGGKLEINSAQEI DRGOYICVATSVAGEKEIKYEVVDVLVPAIEGGDE 1740  
QY 183 ---LL---ASGS-----G-----D-----S-- 191  
Db 1741 TSYFVVMVNNLELDCHVTGSPPTIMWLKQGLIDBERDGFILLNGRLKLVIAQAQVSN 1800  
QY 192 ---TA-----R-----I-W 196  
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPPTIKSSGLSERVVVKYKPVALQCIANGIPNSITW 1860  
QY 197 ---NL-----S 199  
Db 1861 LKDDQPNTAAGNLKIQSSGRVLQIAKTLLEDAGRYTCVATNAAGETQOHIQLHVHEPPS 1920  
QY 200 -E-----N-----S-----T-----SGST----- 207  
Db 1921 LEDAGKMLNLTVLVSNPVQLECAAGNPVPVITWYKDNRLLSGTSMTFLNRQIIDIES 1980  
QY 208 -Q-----L--V-----LR-H-----C-----I-- 215  
Db 1981 AQISDAGIYKCVAINSAGATELFPYSLQVHVAPSISGSNNMVAVVNNPVRLECEARGIPA 2040

QY 216 ---R-----E---G-----G-Q--D--- 221  
Db 2041 PSUTWLKOGSPVSSFNGLQVLGGRIALATSAQISDTRYTCVAVNAAGEKORDIDLRV 2100  
QY 222 -VPSN-----KD-----VT---S-L- 231  
Db 2101 YVPPNMGBEQNVSVLISOAVELLQSDAIPPTLTWLKDGHPLLKPKGLSISENRSVLK 2160  
QY 232 ---D-----W--N--S-----EG--TLL--A 241  
Db 2161 IEDAQVQDTRYTCVATNAGTKERNYNVWPPNIGGSDLTQLTVIBGNLISLICES 2220  
QY 242 T-----GS--Y-D-----G-----F-----A-- 249  
Db 2221 SGIPPPNLTWKKGSPVLTDMSGRVRLISGROLQISIAEKSDAALYSCVASNVAGTAKK 2280  
QY 250 ---R-----I-----WTQD----- 256  
Db 2281 EYNLQVIRPTITNSGSHPTIIVTRGKSISLECEVOGIPPTVTWTKDGHPLIKAKGVE 2340  
QY 257 ---NL-AS-----T---L-----GQH----- 265  
Db 2341 ILDEGHILQKNIHVSQDTRYCVAVNAGMTDKKYDLSVHAPPSIIGNHRSPENISVVE 2400  
QY 266 K-----G--P-I--FALK--W----- 274  
Db 2401 KNSVSLTCEASGIPLSITWF--KDGWPVLSNSVRLISGGRMLRLMQTTMEDAGQYTCV 2458  
QY 275 ---N-----K-K-----GN----- 279  
Db 2459 VRNAAGEERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVEITWHKDGQ 2518  
QY 280 ---F--I-----J-----S---RGV-- 286  
Db 2519 PLOEDBAHHIISGGRFLQITNVQVHTGRYTCVCLASSPAGHKRSFSLNVFVSTIAGVS 2578  
QY 287 D--K-T---T-I-----I-W----- 293  
Db 2579 DGNPDEVTLNSPTSLVCEAYSYPATITWFKDGTPLBSNNRILPGGRTILOILNAQE 2638  
QY 294 D-----A-----H-----TG-----EA 300  
Db 2639 DNAGRYSCVATNEAGEMIKHYEVKYVPIPIINKGLWGFGLSPKPKVKIKVNNLTLECEA 2698  
QY 301 ---K-----K-----Q-----Q-----F 304  
Db 2699 YAIPSASLWYKDGQPLKSDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDELDF 2758  
QY 305 ---P-F-----H----- 307  
Db 2759 DVNIQVPPSFQKLWEIGNMLDTRNGEAKDVIINNPIISLYCETNAAPPPTLTWYKDGHP 2818  
QY 308 -SA-----P-----A-----LDV-----D- 315  
Db 2819 TSSDKVLLPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPIIKGANSDL 2878  
QY 316 ---WO-----SN----- 319  
Db 2879 PEEVTVLNKSALEICLSSGPAPRNSWQDQOPLEDHHPKLSNGRILQILNTOITDI 2938  
QY 320 ---NT-----FAS--C-----S- 326  
Db 2939 GRVYCVAEINTAGSAKKYFNLNVHVPSPVIGPKSENLTVVVNNFISLTCEVSGFPDLSW 2998  
QY 327 ---T-----D--MCI----- 331  
Db 2999 LKNEQPIKLTNTLIVPGGRTLQIIRAKVSDGGEYTCIAINQAGESKKKPSLTVVYPPSI 3058  
QY 332 --H-----V--C-----K-----L-GQ----- 338  
Db 3059 KQHDSESLSVNVNREGTSVLECESNAVPPVITWYKGRMITESTHVILADGQMLHIK 3118  
QY 339 ---D-----R-----P-IK-----TF----- 345



```
Db 3119 KAEVSDTQYVCRAINAGRDDKNFHLNVVPPSIEGPEREVIVETISNPVLTCDATGI 3178
QY 346 -----Q-----G-----H-----T----- 349
Db 3179 PPPTIAWLKNHKKRIENSLSLEVRILSGSKGLQIARSQHSNGYTCIASNMEGKAQKYF 3238
QY 350 -----N-E-V-NA-----IKW--D-P--T----- 360
Db 3239 LSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIQLWKDKGPIASGETERIRVS 3298
QY 361 --G--NLL-A-S-----C-S-----D-----D-----MTL----- 373
Db 3299 ANGSTLNLYGALTSDTGKYTCVATNPAGEBEDRIFENLVNVTPTIRGNKDEAEKMLTLDVT 3358
QY 374 --K-----I-W-----SM-----KO--D--NCV----- 384
Db 3359 SINIECRATGTPPPQINLWLNKGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVTYCVASNR 3418
QY 385 -----H-----D-----L--Q----- 388
Db 3419 AGVDNKHYNLOVPAPNMDNSMCTEITVLKGSSTSMACITDGTAPAPMAWLRDQPLGL 3478
QY 389 -----Q-----H-N-----KEI 394
Db 3479 DAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFILKLVLEPPHINGSHEEII 3538
QY 395 -----YT-----I--K-W----- 399
Db 3539 SVIVNPLELTCIASIPAPKMTWMKGRPLPQTDQVQLGGGEVLRISTAQVEDTGRYT 3598
QY 400 -----SPTG-----P--GT-----N-----N-----N 408
Db 3599 CLASSPAGDDDKKEYLVRVHPNINAGTDEPRDITVLRNQVLTLEKSDAVPPPVITWLRN 3658
QY 409 -----P-----N-----ANLM-LAS--A--S--F----- 420
Db 3659 GERLOATPRVRLISGGRYLQINADLGDANYTCVASNIAGKTRSFILTVNVPNNIKGG 3718
QY 421 -----DSTV-----RL-W-D-V-----DR-----G--I----- 433
Db 3719 PQSLVILLNKSTVLEACIAGVPTPRITWRKOGAVLAGNHARYSILENGFLHQSAHVDT 3778
QY 434 -----C-----I--H-----TL--T----- 439
Db 3779 GRYLCMATNAAGTDRRIDLOVHVPPSIAPGPTMTVIVNVQTLTACEATGPKPSINWR 3838
QY 440 K-H-----Q-E-----P-V-----Y-----S 447
Db 3839 KNGHLLNVQNSYRLSSGSLVLIISPSVDDTATYECTVTNGAGDDKRTVLDLTQVPPS 3898
QY 448 VA-----F-S-----P-D----- 453
Db 3899 IADEPTDFLVTKHAPAVITCTASGVPPPSIHWTKNGIRLLPRGDGYRILSSGAIEIATQ 3958
QY 454 -----GRY-----L--ASG--S-F 462
Db 3959 LNHAGRYTCVARNAGSAHRHVTLHVHPPVIOQPSSELHVILNPIILLPCEATGTPSPF 4018
QY 463 -----D-----KCV----- 466
Db 4019 ITWQKEGINVNTSGRNVHVLPSGGLQISRAVREDAGTYMCVAQNPAQTALGKILNVQVP 4078
QY 467 -----H-----I-W----- 469
Db 4079 PVISPHLKEYVIAVDKPIITLSCADGLPPDPITWTKDGRAIVESIRQVRLSSGLQIAFV 4138
QY 470 -----N-----T-----Q----- 472
Db 4139 QPGDAGHYTCMAANVAGSSSTSKLTVHVPPIRSTEGHYTVNENSQAILPCVADGIPPT 4198
QY 473 -----T-----G-----A-----L-VH- 478
```

## RESULT 13

ADJ70089

ID ADJ70089 standard; protein; 5636 AA.

XX AC ADJ70089;

XX DT 06-MAY-2004 (first entry)

XX DE Human heat mitochondrial protein as a therapeutic target SeqID1895.  
XX KW mitochondrial; human; screening assay; diabetes mellitus;  
XX KW Huntington's disease; osteoarthritis;  
XX KW Leber's hereditary optic neuropathy; LHON;  
XX KW mitochondrial encephalopathy lactic acidosis and stroke; MELAS;  
XX KW myoclonic epilepsy ragged red fibre syndrome; MERRF; cancer;  
XX KW neuroprotective; nootropic; antidiabetic; anticonvulsant; antiarthritic;  
XX KW osteopathic; ophthalmological; cytostatic.

XX OS Homo sapiens.

XX PN WO2003087768-A2.

XX PD 23-OCT-2003.

XX PF 04-APR-2003; 2003WO-US010870.

XX PR 12-APR-2002; 2002US-0372843P.

XX PR 17-JUN-2002; 2002US-038987P.

XX PR 20-SEP-2002; 2002US-0412418P.

XX XX (MITO-) MITOKOR.

XX PA (BUCK-) BUCK INST AGE RES.

XX PI Ghosh SS, Fahy ED, Zhang B, Gibson BW, Taylor SW, Glenn GM;

XX PI Warnock DE;

XX XX WPI; 2003-845369/78.

XX PT Identifying a mitochondrial target for drug screening assays and for  
XX PT treating diseases associated with altered mitochondrial function,  
XX PT comprises detecting a modified polypeptide in a sample and correlating  
XX PT with the disease.

XX PS Claim 1; SEQ ID NO 1895; 180pp; English.

XX CC This invention relates to novel mitochondrial targets that can be used

CC for therapeutic intervention in treating a disease associated with  
 CC altered mitochondrial function. Specifically, it refers to a method for  
 CC identifying proteins of the human heart mitochondrial proteome that are  
 CC useful for drug screening assays, as well as therapeutic targets. The  
 CC present invention describes a method for identifying such proteins that  
 CC can be used in the treatment of various diseases associated with altered  
 CC mitochondrial function including diabetes mellitus, Huntington's disease,  
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON), mitochondrial  
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
 CC compositions have neuroprotective, neurotropic, antidiabetic,  
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
 CC cyostatic activities. This polypeptide sequence is a human heart  
 CC mitochondrial protein of the invention.

XX Sequence 5636 AA;

Query Match 71.5%; Score 2614.8; DB 7; Length 5636;  
 Best Local Similarity 9.5%; Pred. No. 5.2e-31;  
 Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;  
 QY 1 M-S-----I-----SS--D-----EV----- 8  
 Db 1 MISWEVVHTVFLFALLYSSLAQDASPOSEIRAEFFEGASTLAFVFDVTGMYDVLVQVI 60  
 QY 9 -----NF-LV----- 12  
 Db 61 EGASKILETSLKRPKPLFNALVPHDHPGIVPTITDDPKKQYELRELYVQGGDCPE 120  
 QY 13 -----Y-----R--Y-----IQ----- 17  
 Db 121 MSGAIKALEISLPGSFIVFTDARSKDYRLTHEVLQLIQKQSQVVFVLTGDCDDRTH 180  
 QY 18 -----E-----SG--F-----S--H--S-----A-----F 26  
 Db 181 IGYVVEEIASTSGGVFLDKQVNEVLKWVEEAVQAKVHLLSTDLHQAVNTWRIFP 240  
 QY 27 -----T-----I-----FG-----I-----K 31  
 Db 241 DPLKEVTVSLGSPSMIEIRNPLGLKIKGGLHLLNHSKAVNVKPEAGWTVK 300  
 QY 32 ---S--H-----I-----S--Q----- 36  
 Db 301 TSSSGRHSVRIITGLSTIDFRAGFRKPTLDFKKTVSFVQGIPTVYLLNTSGISTPARID 360  
 QY 37 -----S-----NI-----N-----G----- 41  
 Db 361 LLELLSISGSSLKTIPVKYYPHRKPYGIMNISDFVPPNEAFFLKVTGYDKDYLFORVSS 420  
 QY 42 -----A-----L-----VP-----P-----A 47  
 Db 421 VSFSSIVDPAPKVTMPKTEGTYLQPGQIPCSVDSLLPFTLSFVRNGVTGLGVQYLKESA 480  
 QY 48 -----A-----L-----I-----S-----IIQ-----RG----- 56  
 Db 481 SVSLDIKAVTLSDGFFVECIASVAGTGRAQTFDVSSEPPVQVPPNVTVTGERAVLT 540  
 QY 57 -----L--Q-----I-----Y--V--E----- 61  
 Db 541 CLIIISAVDYNLTQWRNDRVRLAEPARTLANLSLELKSFKENDAGEYHCWVSSEGGSS 600  
 QY 62 -A-----EVS-----NE----- 68  
 Db 601 AASVFLTVQBPVKVTMPKQSFSGSEVSMCSATGYPKPKIAWTVMDFIVGSHRYRM 660  
 QY 69 --DGLF-----D-----G-----R--P-----I-- 78  
 Db 661 TSDGTLPIKNAAPKADAGIYGLASNSAGTDKQNSTLRYIEAPKLMVQSELLVALGDIIV 720  
 QY 79 -E--S-----L--S--LI-----DA-----V----- 87  
 Db 721 MECKTSGIPPPQVKWFKGDLRLPSTFLIIDPLLLGLLKIQTQDLQDADGYTCVAINAAGR 780

QY 88 -----M-----P-DV-----VQ-----T----- 94  
 Db 781 ATGKIITLDVSGPPVFIOEPADVSMEIGSNVTLPCYQGYPEPTIKWRLDNMFISRPFS 840  
 QY 95 -----R----- 95  
 Db 841 VSSISQLRTGALFILNLWASDKGTICEAENQFKIQSETTVTVTGLVAPLIGISPSVAN 900  
 QY 96 ---QQ-----A-----Y-----R--D--KL-- 103  
 Db 901 VIEGQQLTPTLLAGNPERRWIKNSAWLQNPYITVRSQGLHIERVQLQDGGGYTC 960  
 QY 104 -----A-----Q--QQ-----AA----- 109  
 Db 961 VASNVAGTNNKTSVVVVHVLPTIQHQQLLSTIEGIPVTLPCASGNPKPSVLSKKGEL 1020  
 QY 110 ---AAA---A-A-----A--AA--A----- 118  
 Db 1021 ISTSAKFSAGADGSLVWVSPEGESGEYVCTATNTAGYAKRQVLTVVYVRPVGDLRG 1080  
 QY 119 -SQ-----O-----CS----- 123  
 Db 1081 LSQDKPVEISVLAGEEVTLPCEVKSUPPIITWAKETQLISPPSRHFLPSSMKITET 1140  
 QY 124 -----A-----K-----N----- 126  
 Db 1141 RTSDSGMVLCVATNIAGNVTOAVKLVNHHVPPKIQRGPKHLKVQVGQVQDIPCAOQTPLP 1200  
 QY 127 -----GE--N--T-----A-----N--G----- 133  
 Db 1201 VITWSKGGSTMLVDBHHVSNPDGTLSDQATPSDAGIVTCVATNIAGTDETEITLHVOE 1260  
 QY 134 -----E-----E--N-----GA-----H----- 139  
 Db 1261 PPTVELEPPYNTTFOERVANQRIEPPCPAKGPKPTIKWLHNGRELTGREPGISILEEG 1320  
 QY 140 T--IA-----N-----N--HT-----DM----- 148  
 Db 1321 TLLVIAVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVKIDKEQVSNVSVLLNQLTNL 1380  
 QY 149 -----M-----E----- 150  
 Db 1381 FCEVEGTSPSIIMWYKDNQVVTESITQTVNNGKILKLFATPEDAGRYSCKAINIAGTS 1440  
 QY 151 ---VD-----G-----DV--E-----I-----P-----PN 160  
 Db 1441 QKYFNIDVLVPPTIIGTNFPKEVSVVLRNDRVALECOVKGTTPPDHFKDGRPLFLGDPN 1500  
 QY 161 -----K-A-----V-----V-----VL 165  
 Db 1501 VELLDRGQVLHLKNARENDKGRYQCTVSNAGKQAKDKIKTIYIPPSIKGGNVTTDISVL 1560  
 QY 166 -----RG-----H----- 168  
 Db 1561 INSLIKLECKTRGLPMPAITWYKDGQPIIMSSSOALYIDKQYLHIIPRAQVSDSATYCHV 1620  
 QY 169 -----E-S-----E-----VFI-----C--A-----W----- 177  
 Db 1621 ANVAGTAKSFHVVDVYVPPMIEGNLATPLNKQVVIASHLTLECNAAAGNPSPILTWLKDV 1680  
 QY 178 -----N-----P-V--SD-- 182  
 Db 1681 PVKANDNPFRIAGGKKLEIMSAQIDRGQYICVATSVAGEKEIKYEVDVLPVPAIEGGDE 1740  
 QY 183 -----LL-----ASGS-----G-----D-----S-- 191  
 Db 1741 TSYFVVMVNNLLELCHVTGSPPTIMLWKGDLIDERDGFKILLNGRKLVIQAQVSNT 1800  
 QY 192 -----TA-----R-----I--W 196  
 Db 1801 GLYRCMAANTAGDHKKEFEVTVHVPPTIKSSGLSERVVVKYKPVALQCIANGIPNSITW 1860  
 QY 197 -----NL-----S 199

Db 1861 LKDDQPNTAQGNLKIQSSGRVLQIAKTLLEDAGRYTCVATNAAGETQOQHQLHVHPPS 1920  
QY 200 -E-----N-----S-----T-----SGST----- 207  
Db 1921 LEDAGKMLNLTVLSNPVQLEKAAAGNPVPIITWYKDNCLLSGSTMTFLNRQOIIDIES 1980  
QY 208 -Q-----L-----V-----LR-H-----C-----I-- 215  
Db 1981 AQISDAGIYKCAINSAGATELFSLQVHVPASISGSSNNMVAVVNNVPVLECEARGIPA 2040  
QY 216 -----R-----E-----G-----G--Q--D-- 221  
Db 2041 PSITWLKDGSPVFSFNGLOVLSGGRILALTSTQISDGRYTCVAVNAAGEKORDILRV 2100  
QY 222 -VPSN-----KD-----VT--S-L- 231  
Db 2101 YVFPNIMEERQNVSLISQAVELLQSDAIPPTLTWLKDGHPLLKPKGLSISENRSVLK 2160  
QY 232 -----D-----W--N--S-----EG--TLL-- 240  
Db 2161 IEAQVQDGTGRYTCATNVAGTEKKNNYVNIWPPNIGGSDELQTLTVIEGNLSILCE 2220  
QY 241 AT-----GS--Y--D-----G-----F-----A- 249  
Db 2221 SSGIPPNLIWKKGSVLTDSMGRVILSGRQLQISIAEKSDAALYSCVASNVAGTAK 2280  
QY 250 -----R-----I-----WTXDG----- 256  
Db 2281 KEYNLQVIRPTITNGSGHPTEIIVTRGKISILECEVOGIPPTVTWMDKGHPLIKAGV 2340  
QY 257 -----NL-AS-----T-----L-----GOH----- 265  
Db 2341 EILDEGHILQKNIHVSDDTGRYCVAVNVAGMTDKYDLSVHAPPSTIGNHRSPENISVV 2400  
QY 266 -K-----G-PI-----FALK-W----- 274  
Db 2401 EKNSVSLTCEASGIPLPSTTFW--KDGWPVLSNSVRILSGGRMLRLMQTTMEDAQYTC 2458  
QY 275 -----N-----K-K-----GN----- 279  
Db 2459 VVRNAGEERKIFGLSVLPVPHIVGENTLEBVKVKEKQSVTLTCEVTGNPVPBITWHKDG 2518  
QY 280 -----F--I-----L-----S--AGV- 286  
Db 2519 QPLQDEAHHIISSGRFLQITNVQVPHGRTYTCCLASSPAGKSRSPSLNVFVSPTIAGVG 2578  
QY 287 -D--K--T-----T--I-----I-W----- 293  
Db 2579 SDGNPEDVTILNSPTSLVCEAYSYPATITWPKDGTPLSNRNRILPGGRTLQILNAQ 2638  
QY 294 -D-----A-----H-----TG-----E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWGLSPKBEVKIKVNNLTLECE 2698  
QY 300 A-----K-----Q--Q----- 303  
Db 2699 AYAI PSASLSWYKDGQPLKSDHVNIAANGHTLQIKEAQISDGRYTCVASNIAGEDEL 2758  
QY 304 F-----P-F-----H----- 307  
Db 2759 FVNIQVPPSFQKLEIGNMLDTGRNKAQVNIINNPISLYCEETNAAPPPTLTWYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315  
Db 2819 LTSSDKVILPGGRVLIQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878  
QY 316 -----WO-----SN----- 319  
Db 2879 LPBEVTVLNKSALIECLSGSPAPRNSWQKQPLLEDHKKFLSNGRILQILNTQITD 2938  
QY 320 -----NT-----FAS--C-----S 326

Db 2939 IGRYVCVAENTAGSAKKYFNLVHVPSPVIGPKSENLTVVVNNFISLTCEVSGFPDPLS 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKXQPIKLTNTLIVPGGRTLOIIRAKVSDGGEYTCIAXNAGESKKKFSLTVYVPPS 3058  
QY 332 -H-----V--C-----K-----L--GQ-- 338  
Db 3059 IKDHDSESLVNVNREGTSVLECESNAVPPVITWYKNGRMITESTHVEILADGQMLHI 3118  
QY 339 -----D-----R-----P-IK----- 343  
Db 3119 KKAEDVTQGYVCRAINVAGRDDKNFNLNVYVPPSIEGPEREVIVETISNPVLTCDATG 3178  
QY 344 -T-----F-----Q--G--H-----T----- 349  
Db 3179 IPPPTIAWLKNYKRIENSLSVRILSGGSKLOIARSQSDSGNYTCIASNMEGKAQY 3238  
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
Db 3239 FLSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTLOIWKDKGPITASGETERIV 3298  
QY 361 -G--NLL-A--S-----C--S-----D-----D-----MTL-- 373  
Db 3299 SANGSTNIYGALTSDTKYTCVATNPAGEEDRIFNLNVYVPTIRGNKDEAEKLTMYD 3358  
QY 374 -K-----I-W-----SM-----KO--D--NCV-- 384  
Db 3359 TSINIECRXTGTPPPQINLKNGLPLPLSSHRLLAAGVIRIVRAQVSDVAVYTCVASN 3418  
QY 385 -----H-----D-----L--Q-- 388  
Db 3419 RAGVDKNHNLQVAFAPPNDMSMGTEITVLKGSSTSMACITDGTAPSMAWLRDGPGLG 3478  
QY 389 -----Q-----H-N-----KE 393  
Db 3479 LDAHLTVTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFLKVLPEPPHNGSEHEE 3538  
QY 394 I-----YT-----I--K--W----- 399  
Db 3539 ISVIVNNPLELTICIASGIPAPKMTWMMKGRPLPQTDQVQTLGGGEVLRISTAQVEDTGRY 3598  
QY 400 -----SPTG-----P--GT-----N----- 407  
Db 3599 TCLASSPAGDDDKXYLVRVHVPNPNIAGTDEPRDITVLRNRQVTLCKSDAVPPVITWLR 3658  
QY 408 N-----P-----N-----ANLM-LAS-A--S--F----- 420  
Db 3659 NGERLOATPRVILSGGRYLQINNADLGTANTYTCVASNIAGKTTREFILTNNVPPNIK 3718  
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433  
Db 3719 GPQSLVILLNKSTVLECIAGVPTPRITWRKOGAVLAGNHARYSILENGFLHIQSAHVT 3778  
QY 434 -----C-----I-----H-----TL-----T----- 439  
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHVPPSIAPGPTMTVIVNVQTLTACEATGPKPSINW 3838  
QY 440 -K--H--Q-E-----P-V--Y----- 446  
Db 3839 RKNHLLNVNDQNSYRLLSSGSLVLIISPSVDDTATYECTVTNAGDDKRTVOLTQVPP 3898  
QY 447 SVA-----F-S-----P--D----- 453  
Db 3899 STADEPTDLVTKHAPAVITCTASGVPPFSIHWKNGIRLLPRGOGYRILSSGAIELAT 3958  
QY 454 -----GRY-----L--S-- 461  
Db 3959 QLNHAGRYTCVARNAGSAHRHVTLLHVHPPVPIQOPQSELHVLNNPILLPCEATGTPSP 4018  
QY 462 F-----D-----KCV----- 466  
Db 4019 FITWQKGINVNTSGRNHVLPSGGIQLISRAVREDAGTYMCVAQNPAQTALGKILNVQV 4078

QY 467 -----H-----I-W----- 469  
 Db 4079 PPVISPLKKEYIAVDKPIITLSCADGLPPDITWHKDKRAIVESIRQVLSGSLQIAF 4138  
 QY 470 -----N-----T-----Q----- 472  
 Db 4139 VQPDAGHYTCMAANVAGSSSTSKLTIVHVPPIRIRSTEGHYTVNENSAQLPCVADGPT 4198  
 QY 473 -----T-----G-----A-----L-VH 478  
 Db 4199 PAINWKKDNVLLANLLGKYTAEPYGELILENVLEDSGFYTCVANNAAGDTHVSLTVH 4258  
 QY 479 -----S-----Y-----R 481  
 Db 4259 VLPTFTLPGDVSLNKGEQLRLSKATGIPLPKLTWTFNNNIIPAHFDSVNGHSELVIER 4318  
 QY 482 -----GT-----G-----IP-----E-----V-C----- 490  
 Db 4319 VSKEDSGTYVCTAENSFGVKATGFVYVKEPPVKGDPYPSNWIEPLGGNAILNCEVKGDP 4378  
 QY 491 -----WN-----A-----AGD----- 496  
 Db 4379 TPTIQNRKGVDBEISHRIQLGSLAIYCTVNEDAGDYTCVATNEAGVVERSMSTLQ 4438  
 QY 497 -----K-V-----A-----SAS----- 503  
 Db 4439 SPPIITLPEVETVINAGKILNCOATGEPQPTITWSRQHSISWDDRVNVLNNSLYIA 4498  
 QY 504 -----D-----GSV-----C-V----- 509  
 Db 4499 DAQKEDTSEFECVARNLWGLVVRVPIVQVHGFGSQSAWRACSVTCGKGIOKRSRLCN 4558  
 QY 510 --L-----DL--R-----K 514  
 Db 4559 QPLPANGGKPCQSGDLEMRNCONK 4582

RESULT 14  
 ADJ83137  
 ID ADJ83137 standard; protein; 5636 AA.  
 XX AC ADJ83137;  
 XX AC  
 XX AC  
 DT 06-MAY-2004 (first entry)  
 XX DE Human hemocentin protein - SEQ ID 128.  
 XX KW NOVX; cytostatic; antiarteriosclerotic; antidiabetic; antiasthmatic;  
 KW antiasthmatic; antiinflammatory; respiratory; antiarthritic;  
 KW dermatological; antibacterial; cerebroprotective; vasotropic; cardiant;  
 KW haemostatic; hypotensive; hepatotropic; neuroprotective; anorectic;  
 KW neurotropic; antitumor; muscular; immunosuppressive; gynaecological;  
 KW antiparkinsonian; anticonvulsant; tranquiliser; osteopathic;  
 KW antinfertility; antipain; cardiomyopathy; atherosclerosis; diabetes;  
 KW cell signal processing; metabolic pathway; asthma; allergy; emphysema;  
 KW autoimmunity; graft-versus-host; arthritis; cancer; stroke; haemophilia;  
 KW obesity; Alzheimer's; pain; chromosome mapping; tissue typing; human;  
 hemocentin.  
 XX OS Homo sapiens.  
 XX PN US2003170630-A1.  
 XX PD 11-SEP-2003.  
 XX PF 21-DEC-2001; 2001US-00032189.  
 XX PR 21-DEC-2000; 2000US-0257495P.  
 XX PR 22-DEC-2000; 2000US-0258171P.  
 XX PR 20-FEB-2001; 2001US-0269940P.  
 XX PR 08-MAR-2001; 2001US-0274192P.

PR 22-MAR-2001; 2001US-0277826P.  
 PR 29-MAR-2001; 2001US-0279840P.  
 PR 11-APR-2001; 2001US-0282981P.  
 PR 13-APR-2001; 2001US-0283656P.  
 PR 31-JUL-2001; 2001US-0309247P.  
 PR 10-AUG-2001; 2001US-0311754P.  
 PR 17-AUG-2001; 2001US-0313331P.  
 XX (ALSO/) ALSOBROOK J P.  
 PA (TCHE/) TCHERNEV V T.  
 PA (LIUX/) LIU X.  
 PA (SPYT/) SPYTEK K A.  
 PA (ZERH/) ZERHUSEN B D.  
 PA (PAT7/) PATTURAJAN M.  
 PA (LEPL/) LEPLEY D M.  
 PA (BURG/) BURGESS C E.  
 PA (SHIM/) SHIMKETS R A.  
 PA (GROS/) GROSSE W M.  
 PA (SZEK/) SZEKERES E S.  
 PA (VERN/) VERNET C A M.  
 PA (LILL/) LI L.  
 PA (CASM/) CASMAN S J.  
 PA (BOLD/) BOLDOG F L.  
 PA (GORM/) GORMAN L.  
 PA (GANG/) GANGOLLI E A.  
 PA (FERN/) FERNANDES E R.  
 PA (RIEG/) RIEGER D K.  
 PA (EDIN/) EDINGER S R.  
 PA (GUNT/) GUNTHER E.  
 PA (MILL/) MILLET I.  
 PA (SCIO/) SCIORE P.  
 PA (ELLE/) ELLERMAN K.  
 PA (MACD/) MACDOUGALL J R.  
 PA (SMIT/) SMITHSON G.  
 XX Alsobrook JP, Tchernev VT, Liu X, Spytek KA, Zerhusen BD;  
 PI Patturajan M, Lepley DM, Burgess CE, Shimkets RA, Grosse WM;  
 PI Szekeres ES, Vernet CAM, Li L, Casman SJ, Boldog FL, Gorman L;  
 PI Gangolli EA, Fernandes ER, Rieger DK, Edinger SR, Gunther E;  
 PI Millet I, Sciore P, Ellerman K, Macdougall JR, Smithson G;  
 XX WPI; 2003-898249/82.  
 DR New NOVX polypeptides and nucleic acid molecules, useful for diagnosing,  
 XX preventing or treating NOVX-associated polypeptide disorder, e.g.  
 XX PT cardiomyopathy, atherosclerosis, diabetes, cancer, Parkinson's disease or  
 XX PT asthma.  
 PS Disclosure; SEQ ID NO 128; 263pp; English.  
 XX The invention relates to a novel isolated NOVX polypeptide. The  
 CC polypeptide demonstrates cytostatic, antiarteriosclerotic, antidiabetic,  
 CC antiasthmatic, antiinflammatory, antiparkinsonian, anticonvulsant,  
 CC antiarthritic, dermatological, antibacterial, cerebroprotective,  
 CC vasotropic, cardiant, haemostatic, hypotensive, hepatotropic,  
 CC neuroprotective, anorectic, neurotropic, antitumor, muscular,  
 CC immunosuppressive, gynaecological, antipruritic, endocrine,  
 CC ophthalmological, osteopathic, antiparkinsonian, anticonvulsant,  
 CC tranquilliser, analgesic, nephrotropic, antinfertility and antipainemic  
 CC activities. The NOVX polypeptide, nucleic acid or antibody of the  
 CC invention may be useful for treating or preventing a NOVX-associated  
 CC disorder, such as cardiomyopathy, atherosclerosis, diabetes or a disorder  
 CC related to cell signal processing and metabolic pathway modulation.  
 CC Furthermore, the NOVX polypeptides may be useful for diagnosing, treating  
 CC or preventing diseases such as asthma, allergies, emphysema, autoimmune  
 CC disease, graft-versus-host disease, arthritis, cancer, stroke,  
 CC haemophilia, obesity, Alzheimer's disease and pain. The nucleic acids may  
 CC be used as hybridisation probes, in chromosome mapping, tissue typing,  
 CC preventive medicine or pharmacogenomics. The current sequence is that of  
 CC a protein of the invention which is related to human NOVX protein.  
 XX Sequence 5636 AA;  
 SQ

Query Match 71.5%; Score 2614.8; DB 7; Length 5636;  
Best Local Similarity 9.5%; Pred. No. 5.2e-31;  
Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S-----I-SS--D---EV-----8  
DB 1 MISWEVHTVFLFALLYSSLAQDASQSEIRABEPPEGASTLAFVFDVTGSMYDDLQVI 60  
QY 9 -----NF-LV-----12  
DB 61 EGASKILETSKRPKPLNFALVPHPDPEIGPVITTTDPKKFYBELRELYVOGGDCPE 120  
QY 13 -----Y---R---Y---LQ-----17  
DB 121 MSIGAIKIALEISLPGSFIVFTDARKDYRLTHEVLQLIQKQSQVFWFLTGDCDDRTH 180  
QY 18 -----E---SG-F-----S-H--S-----A-----F 26  
DB 181 IGYVVEETASTSGQVFLDKQVNEVLKWBEAQAQSKVHLLSTDHLEQAVNTWRIPP 240  
QY 27 -----T-----FG--I-----K 31  
DB 241 DPSLKEVTVSLGSPSMIEIRNPLGKLIKGFGLHLLNHSKVVNVKEPEAGMWTVK 300  
QY 32 --S-H-----I-----S-Q-----36  
DB 301 TSSSGRHSVRITGLSTIDPRAGFRKPTLDFKKTVSFVQGIPTYVLLNTSGISTPARID 360  
QY 37 -----S-----NI-----N-----G-----41  
DB 361 LLELLSIGSSSLKTI PVKYPHPRKPYGIWNISDFVPPNEAFFLKVTGYDKDYLFORVSS 420  
QY 42 -----A-----L---VP-----P-----A 47  
DB 421 VSPSSIVDPAPKVTMEKTPGYLQGOIPCSVDSLLPFTLSFVRNGVTLGVQYKESA 480  
QY 48 -----A---L-----I-----S---IIQ-----KG-----56  
DB 481 SVSLDIKVTLSDEGFECIAVSSAGTGRAQTFDVSERPPVQVNNVTVPGERAVLT 540  
QY 57 -----L-Q-----P-----Y-V--E-----61  
DB 541 CLIIASVDYNTWRNDVRVLAEPARTLANLSLELKSVDKENDAGEYCHMVSSEGGSS 600  
QY 62 -A-----EVS-----NE-----68  
DB 601 AASVFLTVQEPKVTMPKQKSTGGSEVSMCSATGYPKPKIAWTNDMFIVGSHRYM 660  
QY 69 --DGLTF-----D-----G-----R---P-----I-- 78  
DB 661 TSDGTLFIKNAAPKDAIGYCLASNSAGTDKQNSTLYRIEAPKLMVVQSELLVALGDIIV 720  
QY 79 -E--S-----L-S-LI-----DA-----V-----87  
DB 721 MECKTSGIPPPQVWFPGDLAPSTLIIDPLGLLKIQETQDLBAGDYTCVAINAAGR 780  
QY 88 -----M-----P-DV-----VQ-----T-----94  
DB 781 ATGKITLDVGGPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRRLDNMPFISRPFS 840  
QY 95 -----R-----95  
DB 841 VSSISQRTGALFILNWDKTYICEAENQFKIQSETTVTTVTGLVAPLIGISPSVAN 900  
QY 96 ---QQ-----A-----Y-----R---D--KL--103  
DB 901 VIEGQQLTLPCTLLAGNPISRRWIKNSAMLLQNPYITVRSGLSHIERVOLQDGEYTC 960  
QY 104 ---A-----A-----Q-QQ-----AA-----109  
DB 961 VASNAGTNNKTTSSVVVHVLPTIQHQQLLSTIEGIPVTLPCKASGNPKPSVIMSKGEL 1020  
QY 110 ---AAA---A-A-----A-AA--A-----118

DB 1021 ISTSSAKFAGAGDGLYVVSPEGESGEYVCTATNTAGYAKRKVQLTVYVPRVFGDLRG 1080  
QY 119 -SQ-----Q-----GS-----123  
DB 1081 LSQDKPEISVLAGEBVTLPCEVKSLPPPIITWAKETQLISPPSPRHTFLPSSMKITET 1140  
QY 124 -----A-----K-----N-----126  
DB 1141 RTSDSGMVLCVATNIAGNVTAQVKNLVNHPVKIQRGPKHLKVQVGQVVDIPCNAGQTPLP 1200  
QY 127 -----GE--N--T-----A-----N-G-----133  
DB 1201 VITWSKSGTMLVDGEHVSNDPDTLSIDQATPSDAGIYTCVATNIAGTDEBITLHVQE 1260  
QY 134 -----E---E--N-----GA-----H-----139  
DB 1261 PPTVDELEPPYNTTFOERVANORIEPPCPAKGPKPTIKWLHNGRELGTREPGISILEEG 1320  
QY 140 T---IA-----N-----N--HT---DM-----148  
DB 1321 TLLVIASTPYDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKEQVSNVSVLLNQLTNL 1380  
QY 149 -----M-----E-----150  
DB 1381 FCEVEGTSPSIIMWKDQNVQVTTESSTIQTWNNGKILKLFATPEDAGRYSCKAINIAGTS 1440  
QY 151 -----VD-----G-----DV-E-----I-----P-----PN 160  
DB 1441 QKYFNIDVLVPPTIIGTNPFKEVSVVLRDVALECOVKGTPPDHFWKDGKFLGDPN 1500  
QY 161 -----K-A-----V-----VL 165  
DB 1501 VELLDRGVHLKNAARNDKGRYQCTVSNAGQAQDKLTIYIPPSIKGNVTTDISVL 1560  
QY 166 -----RG-----H-----168  
DB 1561 INSLIKLECKTRGLPMPAITWYKQGPIMSSSQALYIDKQYLHIIPRAQVSDSATYCHV 1620  
QY 169 -----E-S-----E-----VFI-----C--A-----W-----177  
DB 1621 ANVAGTAESFHVVDVVPPIEGNLATPLNKQVIAHSITLECNAAAGNPSIITLWKDGV 1680  
QY 178 -----N-----P-V--SD--182  
DB 1681 PVKANDNFRIEAGKKLEIMSAQEI DRGOYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740  
QY 183 -----LL-----ASGS-----G-----D-----S--191  
DB 1741 TSVFIVMNNLLELDCHVTGSPPTIMLWKDGLIDERDGFKILLNGRKLVIQAQVSNV 1800  
QY 192 -----TA-----R-----I-W 196  
DB 1801 GLYRCHAANTAGDHKKBEFTVHVPTIKSSGLSERVVVVKYKVPALQCIANGIPNSITW 1860  
QY 197 -----NL-----S 199  
DB 1861 LKDDQVPTAQNKLKIQSSGRVLIQIAKTLLEDAGRYTCVATNAAGTQOQHILHVHEPPS 1920  
QY 200 -E-----N-----S-----T-----SGST-----207  
DB 1921 LEDAGKMLNETVLVSNPVLCKAAGNPVPVITWYKDNCLLSGTSMTFLNRQIIDIES 1980  
QY 208 -Q-----L--V-----LR-H-----C-----I--215  
DB 1981 AQISDAGIYKCVAINASAGATELFYSLQVHVAPVPSIGSNMVAVVVNNVPRLECEARIPA 2040  
QY 216 -----R-----E--G-----G--Q--D---221  
DB 2041 PSILTWKDGPVSSFNGLQVLSGGRILALSTQISDTRYTCVAVNAAGEKORDILRV 2100  
QY 222 -VPSN-----RD-----VT---S-L-231

Db 2101 YVPPNIMGEQNSVLISQAVELLCSQDAIPPTLTWLDKGHPLKKPGLSISENRRSVLK 2160  
QY 232 -D- - - - -W- -N- -S- - - - -EG- - - - -TLL- - - - - 240  
Db 2161 IEDAQVDTGRYTCEATNVAGTEKKNVNNWPPNIGGSDLTQTUVEGNLSLCE 2220  
QY 241 AT- - - - -GS- -Y- -D- - - - -G- - - - -F- - - - -A- - - - - 249  
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVRLSGRQLOISIAEKSDAALYSVASNVAGTAK 2280  
QY 250 - - - - -R- - - - -I- - - - -WKDG- - - - - 256  
Db 2281 KEYNLQVIRPTITNSGHPTEIIVTRGKSISLECEVQGPPTVTWTKDCHPLIKAKGV 2340  
QY 257 - - - - -NL- -AS- - - - -T- - - - -L- - - - -GOH- - - - - 265  
Db 2341 EILDEGHILQKNLHVSQDTRGYVCVAVNAGMTDKYDLSVHAPPSIIGNHRSPENISVV 2400  
QY 266 -K- - - - -G- -PI- - - - -FALK- - - - -W- - - - - 274  
Db 2401 EKNSVSLTCEASGIPLPSTTW- - - - -KDGMPVSLNSVRILSGRMLRLMQTTMEDAGQYTC 2458  
QY 275 - - - - -N- - - - -K- -K- - - - -GN- - - - - 279  
Db 2459 VVRNAGEERKIFGLSVLPPIHVGENTLEDVKVKERQSVTLTCEVTGNPVEITWHKOG 2518  
QY 280 - - - - -F- -I- - - - -L- - - - -S- - - - -AGV- - - - - 286  
Db 2519 QPLOEDRAHHIISGRFLOITNVQVPHTRYTCCLASSPAGHKRSFSLNVFVPTTAGVG 2578  
QY 287 -D- - - - -K- -T- - - - -I- -W- - - - - 293  
Db 2579 SDGNPEDVTILNSPTSLVCEAYSYPATITWPKDGTPLSNNRILPGGRTLQILNAQ 2638  
QY 294 -D- - - - -A- - - - -H- - - - -TG- - - - -E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWGFLSPKEVKIKVNNLTLECE 2698  
QY 300 A- - - - -K- - - - -Q- - - - -Q- - - - - 303  
Db 2699 AYALPSASLSWYKDGQPLKDDHVNIAANGHTLOIKEAQISDTGRYTCVASNIAGEDEL 2758  
QY 304 F- - - - -P- -F- - - - -H- - - - - 307  
Db 2759 FVNIQVPPSFQKLWEIGNMLDTRNGEAKDVIIINNPISLYCETNAAPPPTLTWYKDGHP 2818  
QY 308 -SA- - - - -P- - - - -A- - - - -LDV- - - - -D 315  
Db 2819 LTSSDKVLIILPGRVLOIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLPPIIEGANS 2878  
QY 316 - - - - -WO- - - - -SN- - - - - 319  
Db 2879 LPBEVTVLNKSALIECLSSGSPAPNSWQKQPLLEDHDKFSLNGRILQILNTQITD 2938  
QY 320 - - - - -NT- - - - -PAS- - - - -C- - - - -S 326  
Db 2939 IGRYVCVAENTAGSAKKYFNLNHVPPSVIPKSENLTVVNNFISLTCEVSGFPPELDS 2998  
QY 327 - - - - -T- - - - -D- - - - -MCI- - - - - 331  
Db 2999 WLKNXQPIKLTNTLIVPGGRTLOIRAKVSDGGEYTCIAXNAGESKKFSLTVVYVPS 3058  
QY 332 -H- - - - -V- - - - -C- - - - -K- - - - -L- -GO- - - - - 338  
Db 3059 IKHDSLSLVNVRECTSVSLECSNAVPPVITWYKNGRMITESTHVEILLADGQMLHI 3118  
QY 339 - - - - -D- - - - -R- - - - -P- -IK- - - - - 343  
Db 3119 KKAESVDTGOVCRAINAGRDKNFHLNVYVPSIEGPEREIVETISNPVTLTCDATG 3178  
QY 344 - - - - -F- - - - -Q- - - - -G- - - - -H- - - - -T- - - - - 349  
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGKLQIARSQSDSGNYTCIASNMEGKAQKY 3238

QY 350 - - - - -N- -E- -V- -NA- - - - -IKW- -D- - -P- - - - -T- - - - - 360  
Db 3239 FLISQVPPSVAGAEIPSDSVLLGENVELVANGIPTLIQWLKDGKPIASGETERIRV 3298  
QY 361 - - - - -G- - - - -NLL- -A- -S- - - - -C- -S- - - - -D- - - - -D- - - - -MTL- - - - - 373  
Db 3299 SANGSLINIYGALTSITGKYTCVATNPAGEEDRIFNLNVVVTPIRGNKDEAKLMTYVD 3358  
QY 374 - - - - -K- - - - -I- -W- - - - -SM- - - - -KO- -D- - - - -NCV- - - - - 384  
Db 3359 TSINIECRXTGTPPPQINMLKNGLPPLSSHIRLLAAGQIVIRVRAQVSDVAVYTCASN 3418  
QY 385 - - - - -H- - - - -D- - - - -L- - - - -L- - - - -Q- - - - - 388  
Db 3419 RAGVDKNHYNLQVAFPPNMDNSMGTEITVLKGSSTSMACITDGTAPASMAWRDGOPLG 3478  
QY 389 - - - - -Q- - - - -Q- - - - -H- -N- - - - -KE 393  
Db 3479 LDAHLLTVSTHGMVLQLLKAFTEDSGKYTCIASNEAGEVSKHFLKLVLEPPHINGSSEHEE 3538  
QY 394 I- - - - -YT- - - - -I- -K- -W- - - - - 399  
Db 3539 ISVINNPLELTCIASGIPAPKMTWMDGRPLPQTDQVOTLGGGEVLRISTAQVEDTRY 3598  
QY 400 - - - - -SPTG- - - - -P- - - - -GT- - - - -N- - - - - 407  
Db 3599 TCLASSPAGDDKEYLVRVHVPPNIAGTDEPRDITVLRNQVTLCKSDAAPPVITWLR 3658  
QY 408 N- - - - -P- - - - -N- - - - -ANLM- -LAS- -A- -S- -F- - - - - 420  
Db 3659 NGERLQATPRVRLSGRYLQINNADLGDANTVTCVASNIAGTTREFILTVNVPNIK 3718  
QY 421 - - - - -DSTV- - - - -RL- -W- -D- -V- - - - -DR- - - - -G- -I- - - - - 433  
Db 3719 GPQSLVILLNKSTVLECIAGVPTPRITWRKDGAVLAGNHARYSILENGLHIQSAHVD 3778  
QY 434 - - - - -C- - - - -I- - - - -H- - - - -TL- - - - - 439  
Db 3779 TGYLCMATNAAGTDERRIDLQVHVPPSIAPGTNTMTVIVNVQTLTACEATGPKPSINW 3838  
QY 440 -K- -H- - - - -Q- -E- - - - -P- -V- - - - -Y- - - - - 446  
Db 3839 RKNHLLNVQDQNSRYLLSSGSLVILSPSVDATYECTVTNGAGDDKRTVDLTQVQVP 3898  
QY 447 SVA- - - - -F- -S- - - - -P- - - - -D- - - - - 453  
Db 3899 STADEPTDFLVTKHAPAVITCTAGVPFPIHWTNGIRLLPRGDGYRILSSGAIFILAT 3958  
QY 454 - - - - -GRY- - - - - 461  
Db 3959 QLNHAGRYTCVARNAGSAHRHVLHVHPPVIOQPSELHVLNNPILLPCEATGTPSP 4018  
QY 462 F- - - - -D- - - - -KC- - - - - 466  
Db 4019 FITWQEGINVTSGRNHVLPSGGLQISRAVREDAGTVMCAQNPAGTALGKILNVQV 4078  
QY 467 - - - - -H- - - - -I- -W- - - - - 469  
Db 4079 PPVISHPLKBYAVDVKPITLSCDAGLPPDPITWHKDGRAIVESIRQVLSGSLQIAF 4138  
QY 470 - - - - -N- - - - -T- - - - -Q- - - - - 472  
Db 4139 VQPGDAGHYTCMAANVAGSSSTSTKLTVHVPPIRSTEGHYTVNENSOAILPCVADGIFT 4198  
QY 473 - - - - -T- - - - -G- - - - -A- - - - -L- -VH 478  
Db 4199 PAINKKNVLLANLLGKYTABPYGELILENVLEDGFPYCVANNAAGEDTHTVSLTVH 4258  
QY 479 - - - - -S- - - - -Y- - - - -R 481  
Db 4259 VLPTTELPDVSINKGBOLRUSCKATGIPKLTWTFNNIIPAHFVSNHSELVIER 4318

QY 482 -----GT-----G-----G-----IP-----E-----V-C----- 490  
 Db 4319 VSKEDSGTYVCTAENSFGVKATGFFVVKPPVFKGDPENWIEPLGNNAILNCEVKGDP 4378  
 QY 491 -----WN-----A-----AGD----- 496  
 Db 4379 TPTIQNRKGVDEIESHRIQLNGSLAIYGTVNEDAGDYTCVATNEAGVVERMSLTQ 4438  
 QY 497 -----K-V-----G-----A-----SAS----- 503  
 Db 4439 SPPIITLEPVETVINAGGKIILNCQATGEPOPTITWSRQHSISWDVRVNLNSLYTA 4498  
 QY 504 -----D-----GSV-----C-V----- 509  
 Db 4499 DAOKEDTSEFCVARNLMGSLVVRVFIQVHGGFQSWAARACSVTCGKGIQKRSRLCN 4558  
 QY 510 --L-----DL--R-----K 514  
 Db 4559 QPLPANGGKPCQSGDLEMRCQNK 4582

## RESULT 15

ADK60205  
 ID ADK60205 standard; protein; 5636 AA.

XX AC ADK60205;

XX DT 06-MAY-2004 (first entry)

XX DE Angiogenesis differentially expressed protein GS-P29.

XX KW vasotropic; antirheumatic; antiarthritic; hypotensive; antiangiinal;  
 KW antiinflammatory; cardiac; angiogenesis inhibitor; gene therapy;  
 KW angiogenesis; endothelial cell; diagnosis; tumor vascularization;  
 KW retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;  
 KW ovary hyperstimulation; psoriasis; endometriosis; restenosis;  
 KW angioplasty; cicatrization; peripheral vascular disease; hypertension;  
 KW vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;  
 KW ischemia; angina; myocardial infarction; chronic heart disease;  
 KW cardiac congestion; macular degeneration; osteoporosis.

XX OS Homo sapiens.

XX XX FR2836687-A1.

XX XX 05-SEP-2003.

XX XX 11-APR-2002; 2002FR-00004546.

XX XX 04-MAR-2002; 2002FR-00002717.

XX XX (GENE-) GENE SIGNAL.

XX XX (ALMA/) AL MAHMOOD S.

XX XX Colin S, Schneider C, Al Mahmood S;

XX XX WPI: 2004-013912/02.

XX XX N-PSDB; ADK60455.

XX XX Compositions for diagnosing, prognosing and treating angiogenic disorders

XX XX including tumor vascularization and heart disease, comprise nucleic acid

XX XX or polypeptide differentially expressed in angiogenesis.

XX XX Claim 7; SEQ ID NO 81; 424pp; French.

XX XX The invention relates to a novel pharmaceutical composition active on

XX XX angiogenesis comprising an endothelial cell nucleic acid whose expression

XX XX is induced by an angiogenic factor and inhibited by an angiostatic agent

XX XX or its complement or fragment, a polypeptide sequence encoded by the

XX XX nucleic acid or its fragment, a molecule capable of inhibiting expression

XX XX of the nucleic acid or a molecule which binds to the polypeptide

XX XX sequence. The invention is used to diagnose, prognose or treat an

XX XX angiogenic disorder in a mammal, particularly a human. The disorder is

CC particularly tumor vascularization, a retinopathy, rheumatoid arthritis,  
 CC Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,  
 CC endometriosis associated with neovascularization, restenosis due to  
 CC angioplasty, overproduction of tissue due to cicatrization, a peripheral  
 CC vascular disease, hypertension, vascular inflammation, Raynaud disease,  
 CC aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,  
 CC myocardial infarction, chronic heart disease, cardiac congestion or  
 CC macular degeneration due to age or osteoporosis. This sequence  
 CC corresponds to a protein encoded by a differentially expressed DNA used  
 CC in the composition of the invention.

XX Sequence 5636 AA;

Query Match 71.5%; Score 2614.8; DB 8; Length 5636;  
 Best Local Similarity 9.5%; Pred. No. 5.2e-31;  
 Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S-----I-----SS---D-----EV----- 8  
 Db 1 MISWEVHTVFLFALLYSSLAQDASQSEIRAEPEEGASTLAFAVFDVTGSMYDDLQVVI 60  
 QY 9 -----NF-LV----- 12  
 Db 61 EGASKILETSLKRPKRPLENFALVPHDPDEIGPVTTITDPKFOYELRELYVGGGDCPE 120  
 QY 13 -----Y-----R---Y-----LQ----- 17  
 Db 121 MSIGAIIKIALEISLPGSFYVFTDARSKDYRLTHEVLQIIQQKQSQVVFVLTGDCDDRTH 180  
 QY 18 -----E-----SG--F-----S--H--S-----A-----F 26  
 Db 181 IGKVVYEETASTSGQVFLDKKQVNEVLKVEEAQVASKVHLLSTDHLEQAVNTWRIFP 240  
 QY 27 -----T----- 31  
 Db 241 DPSLKEVTVSLGSPSPMIEIRNPLGKLIKKGFGLHLLIHNSAKVNVNKEPEAGMWTX 300  
 QY 32 ---S--H-----I-----S--Q----- 36  
 Db 301 TSSSGRHVRITGLSTIDFRAGFSRKPTLDFKTVSRPVQGIPTYYVLLNTSGISTPARID 360  
 QY 37 -----S-----NI-----N-----G----- 41  
 Db 361 LLELLSISGSLKTIPTVKKYYPHKKPYGIWNISDFVPPNEAFFLKVTGYDKDYLFQVSS 420  
 QY 42 -----A-----L-----VP-----P-----A 47  
 Db 421 VSFSSIVPDAPKVTWPKETPGYLPQGPICSDVSLPFTLSFVRNGVTGLGVQYLKESA 480  
 QY 48 -----A-----L-----I-----S-----IIQ-----KG----- 56  
 Db 481 SVSLDIKVTLSDEGFYECIAVSSAGTGRAQTFFDVSEPPVIOVNNVTVTGERAVLT 540  
 QY 57 -----L--Q-----Y-----V--E----- 61  
 Db 541 CLIIISAVDYNLTWQRNDRVRLAEPARTLANLSLELKSVPNDAGEYHCVSSESGSS 600  
 QY 62 -A-----EVS-----NE----- 68  
 Db 601 AASVFLTVQEPKVTVMKPNQSFQSGSEVSMCSATGYPKPKIAMTVNDMFIVGSHRYRM 660  
 QY 69 --DGLTF-----D-----G-----R---P-----I-- 78  
 Db 661 TSDGTLFIKNAAPKAGIYGLASNSAGTDKQNSTLYRIEAPKLMVQSELLVALGDITV 720  
 QY 79 -E---S-----L--S--LI-----DA-----V----- 87  
 Db 721 MECKTSGIPPPQVKKWFGDLRLRPSTFLIIDPLGLGLKIQETQDLQDAGDYTCVAINAGR 780  
 QY 88 -----M-----P-DV-----VQ-----T----- 94  
 Db 781 ATGKITLDVGGPPVFTQEPADVSMETGNSVTLPCYVQGYPEPTIKWRRLDNMPFSRPF 840

95 ---R--- 95  
841 VSSISQRTGALFILNLWASDKGTIYCEAENQFKIOSETTVTGLVAPLIGISPSVAN 900  
96 ---Q--- 103  
901 VIEGQQLTLPCTLLAGNPERRWIKNSAMLLQNPYITVRSGLSHIERVQLQDGGYTC 960  
104 ---A--- 109  
961 VASNVAGTNNKTSVVVHVLFTHQGOQILSTIBGIPVTLFCKASGNPKFSPVIMSKKGL 1020  
110 ---AAA--- 118  
1021 ISTSSAKFAGADGSLYVVSPEGESEYVCTATNAGYAKRVQLTVYVRVFGDLRG 1080  
119 ---SQ--- 123  
1081 LSQDKPVEISVLAGEEVTLPCEVKSLLPPIITWAKETQLISPFSPRHTFLPSSGSKITET 1140  
124 ---A--- 126  
1141 RTSDGMYLCVATNIAGNVTAQVKNLVHVPFKIQRGPKHLKVQGVQDVIPCAQGTPLP 1200  
127 ---GE--- 133  
1201 VITWSKGGTMLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260  
134 ---E--- 139  
1261 PPTVEDLEPPYNTTFQERVANQRIEFPCKAPKTPKTIKWLHNGRELTGREPGISILEEG 1320  
140 T---IA--- 148  
1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKQKEQVSNVSLNQLTNL 1380  
149 ---M--- 150  
1381 FCEVEGTPSIIMYKDNQVTSSTIQTVNNKILKLFRATPEDAGRYSCKAINAGTS 1440  
151 ---VD--- 160  
1441 QKVFNDVLVPPTIIGTFNFKVSNLNRDVALEQVKGTGTPFDIHWFKDGKFLGLDPN 1500  
161 ---K--- 165  
1501 VELLDRQVLHLKNARNDKGRYQCTVSNAGKQAKDIKLTIXIPPSIKGGNVTTDISVL 1560  
166 ---RG--- 168  
1561 INSLIKLECKTRGLPMPAITWYKDGQIPMSSQALYIDKQYLHI PRAQVSDSATYTCV 1620  
169 ---E--- 177  
1621 ANVAGTAEKSFHDVYVPPMIEGNLATPLNKQVVIASHLSLECNAGNPSPIITLWKDGV 1680  
178 ---N--- 182  
1681 PVKANDFRIEAGKKLEIMSAQEI DRGQYICVATS VAGEKEIKEYVDVLVPAIBEGDE 1740  
183 ---LL--- 191  
1741 TSYFIVMNNLELDCHVTGSPPTIMLWKDGLIDERGDKILLNKGRLVIAQAQVNT 1800  
192 ---TA--- 196  
1801 GLYRCMAANTAGHKHEFTVHVPTIKSGLSERVVVVKYKFPALQCIANGIPNPSITW 1860  
197 ---NL--- 199  
1861 LKDDQPNTAGNLIKQSSGRVLIQAKTILLEDAGRYTCVATNAGETQQHIQLHVHPPS 1920  
200 -E---N---S---T---SGST--- 207

1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLSGSTMTFLNRQIIDIES 1980  
208 -Q---L---V--- 215  
1981 AQISDAGIYKCAINAGATFYSLQVHVAPSIGSNMNVAVVWVNNVRLECEARGIPA 2040  
216 ---R--- 221  
2041 PSUTWLKDGSPVSSFNGLQVLGGRIALATSTQISDTRYTCVAVNAAGEKORDIDLRV 2100  
222 -VPSN--- 231  
2101 YVPFNMGBEONVSVLISOAVELLQSDAIPPTTLTWLKDGHLLKKPGLSISENSVLK 2160  
232 ---D--- 240  
2161 IEDAQVQDTRYTCVATNAGTEKKNVNIWVPPNIGGSDELTLQTVIEGNLISLCE 2220  
241 AT---GS---Y---D--- 249  
2221 SSGIPPNLIWKKGSPVLTDMSGRVRIILSGGRQLOISIAEKSDAALYSCVASNAVGTAK 2280  
250 ---R--- 256  
2281 KEYNLQVYIRPTITNSGSHPTIIVTRKKSISLECEVOGIPPTVTWMDKDHPLIKAKGV 2340  
257 ---NL---AS--- 265  
2341 EILDEGHILQLKNIHVSDTRYVCVAVNAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400  
266 -K---G---PI--- 274  
2401 EKNSVLTCEASGIPLPSTTWFP---KDGWPSVLSNSVRILSGGRMLRMQTTMEDAQOYTC 2458  
275 ---N--- 279  
2459 VVRNAGEBERKIFGLSVLPPIHVGENTLEDVKVKEQSVTLTCEVTGNPVPITWKKDG 2518  
280 ---F---I--- 286  
2519 QPLQDEAHIIIGGRFLQITNVQVPHTRYTCVATNAGHTLIQKEAQISDTGRYTCVASNIAGEDLD 2758  
287 -D---K---T--- 293  
2579 SDGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLESNRNIRILPGRTLOILNAQ 2638  
294 -D---A---H--- 299  
2639 EDNAGRYSCVATNAGEMIKHVEKVYIPIINKGDLWGPGLSPKEVKIKVNNLTLECE 2698  
300 A---K--- 303  
2699 AYAI PASLSWYKDGQPLKSDDHVNIAANGHTLIQKEAQISDTGRYTCVASNIAGEDLD 2758  
304 F---P---F--- 307  
2759 FVNIQVPSFQKLWEIGNMLDTRNGEAKDVIIINNPISLYCETNAAPPPTLTWYKDGHP 2818  
308 -SA--- 315  
2819 LTTSSDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANSND 2878  
316 ---W--- 319  
2879 LPBEVTVLNVKALIECLSSGSPAPRNSWQKDGQPLLEDHKKFLSNGRILQILNTQITD 2938  
320 ---NT--- 326  
2939 IGRYVCVAENTAGSAXKYFNLVNHHVPPSVIGPKSENLTVVNNFISLTCEVSGFPFPPDLS 2998  
327 ---T---D---MCI--- 331



Db 2999 WLKXQPIKLTNTLTVPGRTLQIIIRAKVSDGGEYTCIAINXAGSCKKFSLVTVVPPS 3058  
QY 332 ---H---V---C---K---L---GO--- 338  
Db 3059 IKHDSLSLVNVRGTSVLSCESNAPPPVITWYKNGRMITESTHVEILLADGQMLHI 3118  
QY 339 ---D---R---P---IK--- 343  
Db 3119 KKAESDTCQYVCRAINAGRDKNFHLNVVPPSIEGPEREVIVETISNPVTLTCDATG 3178  
QY 344 ---T---F---Q---G---H---T--- 349  
Db 3179 IPPPTIAWLKNYKRIENSOSLEVRILSGGSKLQIARSQSDSGNYTCIASNMEGKAQYY 3238  
QY 350 ---G---NLL-A-S---C-S---D---D---MTL--- 373  
Db 3239 FLSIQVPPSVAGAEIPSDSVLGENVELVCNANGIPTPLIQWLKDGKPIASGETERIRV 3298  
QY 361 ---G---NLL-A-S---C-S---D---D---MTL--- 373  
Db 3299 SANGSTLNIYALTSDTKYKTCVATNPAGEBDRIFNLNVVYVPTIRGNKDEAEKLMYVD 3358  
QY 374 ---K---I-W---SM---KQ-D---NCV--- 384  
Db 3359 TSINIECRXTGTPPPQINMLKNGPLPLSGSHIRLAAGQVIRIVRAQVSDVAVYTCVASN 3418  
QY 385 ---H---D---D---L---Q--- 388  
Db 3419 RAGVDNKHYNLQVAFPNMDSMTGTEITVLKGSSTSMACITDGTAPSMWLDRDQPLG 3478  
QY 389 ---Q---H-N---KE 393  
Db 3479 LDAHVTSTHGMVLQLLKABEDSGKYTCIASNEAGEVSKHFLKVLPEPPHINGSSEHEE 3538  
QY 394 I---YT---I-K-W--- 399  
Db 3539 ISVINNPLELTCIASGIPAPKMTWMDKGRPLPQDQVOTLGGGEVLRISTAQVEDTGRY 3598  
QY 400 ---SPTG---P---GT---N--- 407  
Db 3599 TCLASSPAGDDDKXYLVRVHPVNIAGTDEPRDITVLRNQVTLCKSDAVPPPVITWLR 3658  
QY 408 N---P---N---ANLM-LAS-A-S-F--- 420  
Db 3659 NGERLOATPRVILSGRYLQINNADLGDPTANYTCVASNIAGTKTREFILTNNVPPNKG 3718  
QY 421 ---DSTV---RL-W-D-V---DR---G-I--- 433  
Db 3719 GPOSVLVLLNKSTVLECIAEGVPTPRITWRKOGAVLAGNHARYSILENGFLHIQSAHVTD 3778  
QY 434 ---C---I---H---TL---T--- 439  
Db 3779 TGRYLCMATNAGTDRRRIDLQVHVPSPAPGTNMTVIVNVQTLACEATGPKPSINW 3838  
QY 440 -K-H---Q-E---P-V---Y--- 446  
Db 3839 RKNHLLNVQDQNSYRLLSSGSLVIISPSVDDTATYECTVNGAGDKRTVDLTQVQVP 3898  
QY 447 SVA---F-S---P-D--- 453  
Db 3899 SIADEPTDFLVTKHAPAVITCTASGVFPFPIHWTNGIRLLPRGDGVRILSSGAIBLAT 3958  
QY 454 ---GRY---L---ASG-S- 461  
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIOQOPSELHVILNPNILLPCEATGTPSP 4018  
QY 462 F---D---KCV--- 466  
Db 4019 FITWQKEGINVNTSGRNHVLPSGGLQISRAVREDAGTYMCVAQNAGTALGKILNVQV 4078  
QY 467 ---H---I-W--- 469  
Db 4079 PPVISPHLKEYIADVDPITLSCSADGLPPPDITWHKDGRAIVESIRQVRVLSGSLQIAF 4138

QY 470 ---N---T---Q--- 472  
Db 4139 VQPGDAGHYTCMAANVAGSSSTSTKLTVHVPPIRSTEGHYTWNENSOAILPCVADGPT 4198  
QY 473 ---T---G---A---L-VH 478  
Db 4199 PAINKKONVLLANLIGKYTAEPYBELIENVVLEDSGPYTCVANNAGEDTHVSLTVH 4258  
QY 479 ---S---Y---R 481  
Db 4259 VLPTFTELPDVSLLNKGEOLRLSCKATGIPLKLTTWTFNNIIPAHFDSVNGHSELVIER 4318  
QY 482 ---GT---G---G---IF---E---V-C--- 490  
Db 4319 VSKEDSGTYVCTAENSUGFVKATGFVYVKEPPVFKGDPYPSNMWIEPLGGNAILNCEVKGDP 4378  
QY 491 ---WN---A---AGD--- 496  
Db 4379 TPTIQNRKGVDEISHRIROLGNSLAIYGTVNEDAGDYTCVATNEAGVVERSMSTLQ 4438  
QY 497 ---K-V---G---A---SAS--- 503  
Db 4439 SPPIITLEPVETVINAGGKIILNCQATGEPQPTITWSROGHSISWDDRVNVLNNSLYIA 4498  
QY 504 ---D---GSV---C-V--- 509  
Db 4499 DAQKEDTSEFECVARNLMSGVLVRVPVIVQVHGGFQWSAWRACSVTCGKGIQKRSRLCN 4558  
QY 510 ---L---DL---R---K 514  
Db 4559 QPLPANGGKPCQSGDLEMENCONK 4582

Search completed: January 3, 2005, 15:19:37  
Job time : 119.333 secs

**This Page Blank (uspto)**



Matches 434; Conservative 61; Mismatches 12; Indels 3805; Gaps 385;

1 MS-----SSDE-----V-----N-F-L-----11  
88 MSRLCNGVQCMDSDEGPHCRELQGNCSRLGCHVCQHTLDGPTCYCNSSFQADGKT 147  
12 VY-----R-----Y-LQ-----E-----18  
148 CKDFDESVVGTCSQLCTNTDGSFICCGVEGYLLOPNRSCAKNEPDRPPVILLANSQ 207  
19 -----SG-----FS-----H-SA-----F 26  
208 NILATYLSGAQVSTITPTSTQTTAMDFTSVANETVCWVHVGDSAAQATQKCARMPGLKGF 267  
27 -----TFGIK-S-HISQ-----S-N-----I-----NG-----41  
268 VDEHTINISLHHVEQMAIDWLGTNPFYFVDDIDDRFVCRNRGDTCTVLLDLELYNPKG 327  
42 -AL-----VP-----P-----A-A-48  
328 IALDPAMKVFTDYGOIPKVERCDMDGQNRKLVDSKIVPFHGITLDLYSLRYWADAY 387  
49 -LI-----SIIQK-L-Q-Y-----V-E-----A-EVS-I-----N 67  
388 LDYEVVDYEGKGRQTIQ-GILIEHLYGLTVFENYLYATNSDANAQAQKTSVIRVNRFN 446  
68 -E-----D-GTL-----72  
447 STEYQVTVRVDKGGALHYHORRQPRVRSACENDQYKPGGCDICLLANSHKARTCR 506  
73 -F-----DG-----R-----PIE-----SL-81  
507 RSGFSLGDSKCKPEHELFLYVYKGRPGIIRGMDGAKVPDEHMPIDENLNMPALD 566  
82 -----S-LI-----D-----AV-M-----PD--90  
567 HAETGYTFADTTSYLIGROKIDGTERITLKDGIHNVEGVAVDWMDNLYWTDDGP 626  
91 V-V-----QTR-----Q-----97  
627 ISVARLEKAAQRTKLIEGKMTHPRAIVDPLNGWYMTDWEEDPKDSRRGRLEAWMDG 686  
98 -----A-Y-R-----D-K-----102  
687 SHRDIFVTSKTVLPNGLSLDIPAGRLYWDAFYDRIETILLNGTDRKIYVEGPELNHAF 746  
103 -----L-----AQQAA-----109  
747 GLCHHGYLFWTEYRSGSVYRLRGVGGAPPTVTLRSEPPPIFEIRMYDAQQQVGTNK 806  
110 -----AA-----A-A-----A-----A-RASQ-120  
807 CRVNGGCSLCLATPGSROCAEAOVLADGVTCLANPSYVPPQCPQGEFACANSRC 866  
121 -Q-----G-----S-----A-----K-----N-G-----E 128  
867 IQERWKCDGNDCLDNDSEAPALCHQHTCPDRFKENNRNCIPNRWLCDGNDGSEDE 926  
129 -N-T--A-----N-----G-----E-----E--N 136  
927 SNATCSARTCPFPNQFSCASGRICIPISWTCDDDDCGDRSDESASCAYPCTCPLTQFTCNN 986  
137 G-----A-H-----T-----I-----141  
987 GRGININWRCDNDGNDSEAGCSHSCSSTOFKCNHGRGCIPEHWTCDGNDGDSDE 1046  
142 --AN--N-----HTD-----M-----M-----E--V-D-152  
1047 THANCINQATPPGGCHTDFQCRLDGLCIPLRWRCDGDTDCMDSSDEKSCGTHVCDP 1106  
153 -G-D-----V-----E-I--PP-----NKAV-----V-164  
1107 SVKFGCKDSARCISKAWVCDGNDGNDCEDNDSDENCESLACRPPSHPCANNSTSVCLPDKLC 1166

QY 165 -----L-R-----G-H-----E-S-----E-----V--172  
Db 1167 DGNDGCGDSDEGELCDQCCLNNGGCSHNCVAPGEGIVCSCLPMELGPDNHTCQIQSY 1226  
QY 173 -----F-I-CA-----W-----N-----178  
Db 1227 CAKHLKCSQKQDNKFSVKSCYEGWVLEPBGDSCHSLDPFKPFIIFSNRHRIRRIDLHK 1286  
QY 179 -----P-----V-----S-----D-----LL--A-S-----186  
Db 1287 GDYSLVPLGLRNTIALDFHLSQALYWTVDVEDKIYRGKLLDNGALTSPFEVVIQYGLATP 1346  
QY 187 -----G-----S-----GD-----S-----T--192  
Db 1347 EGLAVDWIAGNIYVESNLDOIEVAKLDGTLRTLLAGDIEHPRALDPRDGLFTWDW 1406  
QY 193 -A--RI-----W-N-----197  
Db 1407 DASLPRIEASMSGAGRRTVHRETSGGWPNGLTVDYLEKRIWLWDARSDAIYSARYDGS 1466  
QY 198 -----LS-----E-----NST-----SG-----S-TQ--L 209  
Db 1467 GHMEVLRGHEFLSHPPFAVTLYGVEVYTDWRTN-TLAKANKMTGHNVTVVQRTINTOPFDL 1525  
QY 210 -V-----L--R-----H--C-----214  
Db 1526 QVTHPSRQWPNPCBANGGQPCSHLCLINYNRTVSCACPHLMKHLKONTTCYEFKKFL 1585  
QY 215 -----I-----RE-----217  
Db 1586 LYARQMEIRGVLDLAPYVNIISFTVPDIDNVTVDYDAREQVYWSVDRVTOAIKRAFIN 1645  
QY 218 G-G-----QDVPSNK-----D-V-----TS-----LD-----232  
Db 1646 GTGVETVWSADLP-NAHGLAVDWVSRNLFWTSYDTNKKQINVARLDGSKFNAVVOGLEQP 1704  
QY 233 -----W--N-S-----EG-----TLL--A--TG-----S-Y-----245  
Db 1705 HGLVHPLRGKLYWTGDNI SMANMDGNSNRTLLFSQKQPGVGLAIDFPESKLYWISSGNH 1764  
QY 246 -----D-GF--A-R-----I-----W-----T--K-DG-----256  
Db 1765 TINRCLDGSGLVIDAMRSQKATALAIMGDKLWADQVSEKMGTCADGSGSVLRL 1824  
QY 257 N--LA-----S-----T-----L--GO-264  
Db 1825 NSTTLYMHMKVDESIOLDHKGNTNPNCSVNNGDCSQLCPTSETTRSCMCTAGYSLRSGQ 1884  
QY 265 -----HKG-----P-----I-F-A-----L-272  
Db 1885 ACEGVCSFLYSVHEGIRGIPDPNDKSDALVPVSGTSLAVGIDFHAENDTIYVWDMGLS 1944  
QY 273 -----K--W-----N--K-----KGN-----F-I--L-----S 283  
Db 1945 TISRAKDQTRWDVVNTNGIGRVEGIADWIAAGNIYWTGQGFVIEARLNGSFYVVIS 2004  
QY 284 AGVDK-----T-----T-----T-----T-----292  
Db 2005 QGLDKPRAITVHPEKGYLFWTEWGOYPIERSRLDGTERTVVLNVNLSWPNGISVDYQDG 2064  
QY 293 --W-DA-----H-----296  
Db 2065 KLYWCARDTKTERIDLETGENREVVLSNNMDMFSVPEDFLYWSRTHANGSIKRG 2124  
QY 297 -----TG-----298  
Db 2125 KDNATDSVPLRTIGIVQLKDIKVFNRDRQKGTNVCAVANGCCQOLCLYRGRGORACAH 2184  
QY 299 -----E-A-----K-----Q-----Q-----F--304  
Db 2185 GMLAEDGASCREYAGYLLYSERTILKSIHLSERNLNAPVQPPEDPEHMKVIALAFDYR 2244

QY 305 -----P-F-H-----S-A----- 309  
Db 2245 AGTSPGTPNRIFPSDIHFQNIQIINDGSRITIVENVGSVEGLAHRGWDTLWTSYTT 2304  
QY 310 -----P-AL-----D-V-D-----W-Q-----S 318  
Db 2305 STITRHTVQTRGAFERETVITMSGDHPRAFLVDECONLMFWTNWNEQHPSIMRAALS 2364  
QY 319 -N-----N-----T----- 321  
Db 2365 GANVLTLIEKDIRTPNGLAIDHRAEKLYFSDATLDKIERCEYDGSRYVILKSEPVHPFG 2424  
QY 322 -----F-----A-----SC-S----- 326  
Db 2425 LAVYGEHIFWTDWVRRAVQRANKHVGSMKLLRVDIPQOPMGIIAIVANDTNSCELSPCRI 2484  
QY 327 -----T-----DM-----CI----- 331  
Db 2485 NNGGQDLCLLTHQHVNCSCRRGRILODLDLTCRAVNSCRAQDEPECANGECINFSLTC 2544  
QY 332 -H-----V-----C-K----- 335  
Db 2545 DGVPHCKDSDEKPSYCNRRCKTFRQCSNGRCVSNMLWCNAGDDCGDGSDBIPCNKTA 2604  
QY 336 -----LG-----Q----- 338  
Db 2605 CGVGEPRCRDGTICGNSSRCNQFVDCEDASDENMCSATDCSSYFRLGVGVLPQPCERTS 2664  
QY 339 -----D-RP-----I-----K----- 343  
Db 2665 LCYAPSVWCDGANDCGDYSDERDCPGVGRPCPLNYFACPSGRCI PMSWTCKDEDDCEHG 2724  
QY 344 -T-----F-----O-----G-H-----T 349  
Db 2725 EBETHCNKFCSEAQFQCNHRCISKOWLCDGSDDCGSDGDEAAHCBGKTCGPSSFCPGT 2784  
QY 350 -----N-----EV-N-A-I-K----- 356  
Db 2785 HVCVPERWLCDDKADGADESIAAGCLYNSTCDDREFWQNRQCI PKHFVCDHBRDCA 2844  
QY 357 -----W-----D-P-T----- 360  
Db 2845 DGSDESPECEYPTGPFSEFRANGRLSSRQWECGENDCHDQSDAEPKNPHCTSPHRC 2904  
QY 361 -----G-N-----L----- 363  
Db 2905 NASSQFLCSSGRVABALLCNGQDCCGSDSDERGCHINECLSRKLSGCQDCEDLKIGFK 2964  
QY 364 -L-----A-S-CS-----D----- 369  
Db 2965 CRCRPGRLKDDGRTCADVDECSSTTPPCSORCINTHSYKCLVEGYAPRGDPHPSCKAV 3024  
QY 370 -D-----M-----T-LK-----I-W-----SM-- 378  
Db 3025 TDEEPFLIFANRYLRKLNLDGNSYTLKQGLNNAVALDFYREQMIYMTDVTTOGSMIR 3084  
QY 379 -K-----Q-----D-N-C----- 383  
Db 3085 RMHLGNSVQVLHRTGLSNPDGLAVDWVGNLYWCDKGRDTIEVSKLNGAYRTVLVSSGL 3144  
QY 384 -----V-----H----- 385  
Db 3145 REPRALVVDQNGLYWTDWGDHSLGRIGMDSSRSVIVDTKITWPNGLTLDYVTERIY 3204  
QY 386 -----D-----L-Q-----Q-H-----N-K----- 392  
Db 3205 WADAREDIIEFASLDGNSRHRVLSQDIPHIFALTFLFEDYVYVTDWETKSNRAHKTGTN 3264  
QY 393 E-I-----Y-- 395  
Db 3265 KTLIISTLHPMDLHVHALRPQDPVPHPCVKNVNGCSNCLLSPGGHKACPTNFYLG 3324  
QY 396 -----T-----IK-W-----S-----P-----TG 403

RESULT 2

Db 3325 SDGRTCVSNCTASQFVKNDKCIPIFWMKDTEDDCGHSDDEPPDCPEFKRPGQFCSTG 3384  
QY 404 -----P-----G-----TN-----N----- 408  
Db 3385 ICTNPAPICDGNDCQDNDSDEANCDIHVCLPSQFKCTNTNRCIPGIFRCNQDNCGDGED 3444  
QY 409 -----PN-----ANL-M----- 414  
Db 3445 ERDCPEVTCAPNQFQCSITKRCIPRVWVCDRNDNDVDSDEPANCTQMTGVDPEFRCKDS 3504  
QY 415 -L-A----- 416  
Db 3505 GRICIPARWKCDGEDDCGSDGDBEPKECBRTCEPYQFRCKNNRCVPRWQCDYDNDCGDN 3564  
QY 417 -----S-----A-S-----F----- 420  
Db 3565 SDEESCTPRPSESESEFSCANGRCIAORWKCDGDHDCADGSDKDCPTPCMDMOFQCKSGH 3624  
QY 421 -----D-S-----T-VR-----L-W-----D 428  
Db 3625 CIPLRWCDADADCMGSDDEEACGTGVRTCLPDEFQCNNTLCKPLAWKCDGEDDCGNSD 3684  
QY 429 -----V-----DRGIC-I-----HTLTKH- 441  
Db 3685 ENPEECARFVCPNRPFRCKNDR-VCLWIGRQCDGTDNCGDGTDEEDCEPPTAHT-T-HC 3741  
QY 442 -----Q-----E-P-----V-- 445  
Db 3742 KDKKEFLCRNQLGSSLRNMFDDCGSDGDEEDCSIDPKLTSCATNASICGDEARCVRT 3801  
QY 446 -Y-----S-V-----A-F----- 450  
Db 3802 EXAAYACASGFTVPGQPCQDINECLRFGTCSQLCNNTKGHLCSARNFMKTHNTCK 3861  
QY 451 -----S-P-----D-GR----- 455  
Db 3862 ABGSEYQVLYIADNEIRSLFPGHPSAYEQAGQDSVRIDAMDVHVKAGRVYTNWHT 3921  
QY 456 -Y-L-A-S-----G-S-----F-----D- 463  
Db 3922 GTISYRSLPAPPTTSNRHRRQIDRGVTHLNI SGLKMRPGIAIDWVAGNVYWTDSGRDV 3981  
QY 464 -K-C-----V-H-I-----W-N-----T-- 471  
Db 3982 IEVAQMKGENRKTLSGMIDEPHAIWVDPRLGTWYMSDGNHHPKIETAAAMDGLRETLVQ 4041  
QY 472 -Q-TGAL-V-HS-Y-----R-GT-----G----- 484  
Db 4042 DNIQWPTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPIVAADSKRGLSHPSIDVFED 4100  
QY 485 -G-----I-F-----EV-----C-W-- 491  
Db 4101 YIYGVTYINNRVFKHKFGHSPVLNLTGGLSHASDVLYHQHKQPEVTNPNCDKCKCEWLC 4160  
QY 492 -----N-----A-----A----- 494  
Db 4161 LLSPSGPVCTCPGKGLDNGTCVVPSPPTPPDAPRPGTCNLOCFNGGSCFLNARQPKC 4220  
QY 495 -----GDK-----V-G-A-- 500  
Db 4221 RCQPRYTGDKCELDQCEWHECRNGGTCAASPSGMPTRCPCPTGTPKTCQVCAGYCANNS 4280  
QY 501 -----S-----ASDGS----- 506  
Db 4281 TCTVNGNQPOQRCLPGFLGDRQYRQCSYCFNFCQMAADGSRQCRCTAYFEGRCE 4340  
QY 507 V-----CVL-----DL-----R-----K 514  
Db 4341 VNKCSRLEGACVWVKQSGDVTNCNCTDGRVAPASCLTCVGHCSNGSGCTWNSK 4392

US-08-469-658-52	Query Match	70.5%; Score 2580.5; DB 2; Length 4544;
Sequence 52, Application US/08469658	Best Local Similarity	10.1%; Pred. No. 2.4e-32;
Patent No. 5917018	Matches 434; Conservative	61; Mismatches 12; Indels 3805; Gaps 385;
GENERAL INFORMATION:		
APPLICANT: Th egeresen, Hans Christian	1 MS	SSDE
APPLICANT: Holtet, Thor Las	88 MSRLNGVQDMSDEGPHCRELOGNSRLCCQHCVP	TLTDLGPTCYCNSSFQADGKT
APPLICANT: Ezerodt, Michael	12	Y-LQ
TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOLDING OF	148 CKDFDECSVVGTCQLCTNDGDSFICGCVGELLQPNRSCAKNEPVDPRPVLIIANSQ	207
TITLE OF INVENTION: PROTEINS	19	SG
NUMBER OF SEQUENCES: 58	208 NILATVLSGAQVSTTPTSTROTMTAMDSYANETVCWVHVGDSSAQTLKCARMPGLKGF	267
CORRESPONDENCE ADDRESS:	27	TEGIG-S-HISQ
ADDRESSEE: Fish & Richardson P.C.	268 VDEHTINISLSLHHVEQMAIDLWTCNFYVDDIDDRIFVCNRNGTCTVTLTDLLELYNPKG	327
STREET: 225 Franklin Street	42	AL
CITY: Boston	328 IALDPAMGVFFTDYQIPKVERCDMDGQNRKLVDSKIVFPFHGITLVLRSVLYWADAY	387
STATE: Massachusetts	49	LI
COUNTRY: USA		
ZIP: 02110-2804		
COMPUTER TYPE: Floppy disk		
MEDIUM TYPE: PC-DOS/MS-DOS		
OPERATING SYSTEM: PC-DOS/MS-DOS		
SOFTWARE: Patent In Release #1.0, Version		
SOFTWARE: #1.25		
CURRENT APPLICATION DATA:		
APPLICATION NUMBER: US/08/469,658		
FILING DATE: June 5, 1995		
CLASSIFICATION: 530		
PRIOR APPLICATION DATA:		
APPLICATION NUMBER: 08/192,060		
FILING DATE: February 4, 1994		
CLASSIFICATION: 530		
ATTORNEY/AGENT INFORMATION:		
NAME: Paul T. Clark		
REGISTRATION NUMBER: 30,162		
REFERENCE/DOCKET NUMBER: 06363/002002		
TELEPHONE: 617 542 5070		
TELEFAX: 617 542 8906		
TELEX: 200154		
INFORMATION FOR SEQ ID NO: 52:		
SEQUENCE CHARACTERISTICS:		
LENGTH: 4544 amino acids		
TYPE: amino acid		
STRANDEDNESS:		
TOPOLOGY: linear		
MOLECULE TYPE: protein		
US-08-469-658-52		

Db 1467 GHMEVLRGHEFLSHPPFAVTLTYGGSEVYWDWRTN-TLAKANKTGHNVTVVQRTNTPQFDL 1525  
Qy 210 -V-----L---R-----H---C-----214  
Db 1526 QVYHPSROPWPNPCEANGGPGCSHLCLINYNRTYSCACPHLMKHLKONTTCYEFKKFL 1585  
Qy 215 -----I-----RE-----217  
Db 1586 LYARQMEIRGVDLDAPYNYIISFTVPDIDNVTLVDYDAREQRYWSDVTRQAIKRAFIN 1645  
Qy 218 G-G-----ODVPSNK-----D-V-----TS-----LD-----232  
Db 1646 GTGVETVVSADLP-NAHGLAVDWVRNLFWTSYDTNKKQINVARLDGFSKNAVVOGLEQP 1704  
Qy 233 -----W-----N-S-----EG---TL---A---TG-----S-Y-----245  
Db 1705 HGLVHPLRCKLYWTDGDNISMANNMGDSNRTLLFSGQKGPVGLAIDFPESKLYWISGNH 1764  
Qy 246 -D-GF---A-R-----I---W-----T---K-DG-----256  
Db 1765 TINRCLDGSGLVIDAMRSLQKATALAIMGDKLWADQVSEKMGTCCKADGSGSVLR 1824  
Qy 257 N---LA-----HKG-----P-----I-F-A-----L-272  
Db 1825 NSTTLVHMKVYDESIQLDHKGTNPCSVNNGDCSLCLPTSETTRSCMCTAGYSLRSGQ 1884  
Qy 265 -----K-----W-----N-----KGN-----F-I---L-----S 283  
Db 1885 ACEGVGSFLLYSVEHIGRIPDPNDKSDALVPVSGTSLAVGIDFHAENDTIYVNDMGLS 1944  
Qy 273 -K---W-----N-----K-----KGN-----F-I---L-----S 283  
Db 1945 TISRAKDQRTWEDVNTGIRVEGIAVDIAGNIYWDQGFVIEVARLNGSFRYVVIS 2004  
Qy 284 AGYDK-----T-----T-----I-----I-----292  
Db 2005 QGLDKPRAITVHPEKGYLFWTEWQYPIERSRLDGTERTVYLVNVSISPNFNGISVDYQDG 2064  
Qy 293 -W-DA-----TG-----K-----Q-----Q-----F-----298  
Db 2065 KLYWCDAKDARTKIBRIDLETGENREVLSNNMDMFSVSPEDFIYWSDRTHANGSIKRG 2124  
Qy 297 -----TG-----K-----Q-----Q-----F-----304  
Db 2125 KDNATDSVPLRTGIGVQLKDIKVPNRDROKGTNVCAVANGGCOQLCYRGRORACACAH 2184  
Qy 299 -----E-A-----K-----Q-----Q-----F-----304  
Db 2185 GMLAEDGASCREYAGYLLYSERTILKSIHLSDERNLNAVPQPPEDPEHMKNVIALAFDYR 2244  
Qy 305 -----P-----E-H-----H-----S-----A-----309  
Db 2245 AGTSPGTNRIFFSDIHFGNIQIINDGSRRIIVENVGSVEGLAVHGWDTLYWTSYTT 2304  
Qy 310 -----P-AL-----D-----V-D-----W-Q-----S 318  
Db 2305 STITRHTVQTRPGAFERETVITMSGDDHPAPFVLDECNLMFWTNWNEQHPSIMRAALS 2364  
Qy 319 -N-----N-----T-----T-----321  
Db 2365 GANVLTLIENDITPGLAIDHRAEKLIFYSDATLDKIERCEYDGSYRILKSEPVHPFG 2424  
Qy 322 -----F-----P-----A-----SC-S-----326  
Db 2425 LAVYGEHIFWTDVVRRAVQANKHVSNNKLLRVDIPQPMGIIVANDTNSCELSPCRI 2484  
Qy 327 -----T-----DM-----CI-----331  
Db 2485 NNGGCQDCLLTHQGHVNCRCGRILQDDLTCTRAVNSSCRAQDEFECAINGEINFSLTC 2544  
Qy 332 -H-----H-----V-----C-K--335  
Db 2545 DGVPCHKDKSDEKPSYCNRRCKKTFRCQSNRCVSNMLWCMGADDCGSDGSDBIPCNKTA 2604

Qy 336 -----LG-----Q-----338  
Db 2605 CGVGEFRCDGTGICGNSSRCNQFVDCEDASDENMCSATDCSSYFRLGVKGVLPQCBRTS 2664  
Qy 339 -----D-----RP-----I-----K-----343  
Db 2665 LCYAPSVWCDGANDCGYSDERDCPGVGRPCPLNYFACPSGRCIPMSWTCDEDDCEHG 2724  
Qy 344 -T-----F-----Q-----G-----H-----T 349  
Db 2725 EDETHCNKFCSEAFQECQNHRCISKQWLCDGSDGSDGDEAAHCEGKTCGPSFSFCPGT 2784  
Qy 350 -----N-----EV---N-A-I-K-----356  
Db 2785 HVCVPERWLCDDKOCADGADESIAAGCLYNSTCDDREFMCQNRQCIKPHFVCDHORDCA 2844  
Qy 357 -----W-----D-----P-T-----360  
Db 2845 DGSDESPECEYPTGSPSEFRANGRLSSRQWECDEGENDCHQSDBAKPNPHCTSPHKC 2904  
Qy 361 -----G---N-----L-----363  
Db 2905 NASSQFLCSSGRCAEALLCNGODDCGSDSDERGCHINECLSKLSGCSQDCBDLKIGFK 2964  
Qy 364 -----L-----A-----S-----CS-----D-----369  
Db 2965 CRCRPGRLKDDGRTCADVDECSSTTFPCSQRCINTHGSYKCLCVEGYAPRGGDPHSCAV 3024  
Qy 370 -D-----M-----T-LK-----I-W-----SM--378  
Db 3025 TDEEPFLIFANRYLRKLNLDGNSYLLKQGLNNAVALDFVREOMIYMTDVTTOGSMIR 3084  
Qy 379 -K-----Q-----D---N---C-----383  
Db 3085 RMHLNSNYQVLRHTGLSNPDGLAVDWGGLNYCWDKGRDTTEVSKLNGAYRTVLVSSGL 3144  
Qy 384 -----V-----H-----385  
Db 3145 REPRALVVDVQNGLYWTDWGDHSLIGMDGSSRSVIVDTKITWPNGLTLDYVTERIY 3204  
Qy 386 -----D-----L-----Q-----Q-----H-----N---K-----392  
Db 3205 WADAREDIYEFASLDGSRHVVLSQDIPHFALTLPEDYVYWTDMETKSNRAHKTGTN 3264  
Qy 393 E---I-----Y---395  
Db 3265 KTLISTLHRPMDLHVHFAHQDPVPHPCVKNVNGCSNCLLSPGGHKKACPTNFYLG 3324  
Qy 396 -----T-----IK-W-----S-----P-----TG 403  
Db 3325 SDGRTCVSNCTASQFVCKNDKCLIPFWWKCDTEDDCGDHSDEPPDCPEFKCRPQFQCGSTG 3384  
Qy 404 -P-----G-----TN-----N-----408  
Db 3385 ICTNPAPICDGNDCQDQNSDEANCDIHVCLPSQFKCTNTRCIPGIFRCNQDNCGDGED 3444  
Qy 409 -----PN-----ANL---M-----414  
Db 3445 ERDCPEVTCAPNQFQCSITKRCIPRVWVCDRNDVCGSDSEFANCTQMTCGVDEFRCKDS 3504  
Qy 415 ---L-A-----416  
Db 3505 GRICIPARWKCDGEDDCGSDSEPFKECDERTCEPYQFRCKNNRCVPRWQCDYDNDGDN 3564  
Qy 417 -----S-----A---S-----F-----420  
Db 3565 SDEESCTPRPCSESEBFCANGRCIAGRWKCDGHDGADGSDERKDTPCRCMDQFQCKSGH 3624  
Qy 421 -----D-S-----T-VR-----L-W-----D 428  
Db 3625 CIPLRWRCADADCMGDSDEEACGTGVRTCPLEFQCNNTLCKPLAWKCDGEDDCGSDNSD 3684

QY 429 -----V-----DRGIC-I-----HTLTKH- 441  
Db 3685 ENPECARFCVCPNRPCKNDR-VCLWIGROCDGTNCGDGTDEDCBPPTAHT-T-HC 3741  
QY 442 -----Q-----E-----P-----V-- 445  
Db 3742 KDKKEFLCNRQCLSSSLRCNMFDDCGDGEDDCSIDPKLTSCATNASICGBEARCVRT 3801  
QY 446 ---Y---S---V-----A-F----- 450  
Db 3802 EKAAYCACRSRGFTVPGQPGCODINECLRFCTCQLCNNTKGHLCSARNFMKTHNTCK 3861  
QY 451 -----S-----P-----D-----GR----- 455  
Db 3862 AEGSEYQVLYIADNEIRSLFPGHPSAYEQAFQGESVRIADMDVHVKAGRVYTNWHT 3921  
QY 456 ---Y-L-A-S-----G-----S-----F-----D- 463  
Db 3922 GTISYRSLPPAAPPTTSNRHRRQIDRGVTHLNISGLKMPRGIALDWAGNVYVTDSDROV 3981  
QY 464 ---K-C-----V-H-I-----W-N-----T--- 471  
Db 3982 IEVAQMKGENRKTLLISGMIDEPHAIVVDPLRGTMWSDMGNHPKRIETAAMDGLRETLVQ 4041  
QY 472 ---Q-TGAL-V-HS---Y-----R-GT-----G----- 484  
Db 4042 DNTQWPTG-LADVHNERLYWADAKLVIGSIRLINGTDPPIVAADSKRGLSHPSIDVFED 4100  
QY 485 ---G-----I-F-----N-----A----- 494  
Db 4101 YIYGVTYINNVRFKHKFGHSLVNLTLGSLHSDVVLVYHQHQPEVNTPCDRKKCEWLC 4160  
QY 492 -----N-----A----- 494  
Db 4161 LLSPSGVCPCPNKRLDNGTCVVPSPPTPPDAPREGTCNLCQFNCGSCFLNARRQPKC 4220  
QY 495 -----GDK-----V-G-A-- 500  
Db 4221 RCPRYTGDKCQLDQWEHCNNGTCAASPSGMPTRCPTGPTGPKCTQVCAGYCANNS 4280  
QY 501 -----S-----ASDGS----- 506  
Db 4281 TCTVNOGNQPCRLPGLGDRQYRCQSGYCNFGTCQWADGSRQCRCTAVFEGSRCE 4340  
QY 507 V-----CVL-----DL-----R-----K 514  
Db 4341 VNKSRCLGACVYVVKOSGVDTCNCTDGRVAPSLCTCVGHCSNGSGCTMNSK 4392

## RESULT 3

US-09-334-220-2  
; Sequence 2, Application US/09334220  
; Patent No. 6323177  
; GENERAL INFORMATION:  
; APPLICANT: St. Jude's Children's Research Hospital  
; APPLICANT: Curran, Thomas  
; APPLICANT: D'Arcangelo, Gabriella  
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
; TITLE OF INVENTION: THERAPIES  
; FILE REFERENCE: 2427/0F704  
; CURRENT APPLICATION NUMBER: US/09/334,220  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 3461  
; TYPE: PRT  
; ORGANISM: Mus musculus  
US-09-334-220-2

Query Match 70.3%; Score 2572.8; DB 3; Length 3461;  
Best Local Similarity 11.8%; Pred. No. 7.7e-33;

Matches 395; Conservative 90; Mismatches 16; Indels 2842; Gaps 361;  
QY 1 M-----S-----S-----S-----S-----D-----EV----- 8  
Db 1 MERGCWAPRALVLAIVLLATLARAATGYYPFPFFLCTHHGBLEGDEGEVLISL 60  
QY 9 ---N-----F-LV--YRY--LQ--ES--CF--S-H 23  
Db 61 HIAGNTYVYVGOEYHVTISTSTFFDGLLVTLG-YTSTSIQSSQSIGSSAFGFGIMSDH 119  
QY 24 -----S-A-----F-----T-----F-----G-I-K----- 31  
Db 120 QFGNQFMCWSVASHVSHLPTNLSPVFIAPAGTGCNFMATATHRGOVIFKDALAQOLC 179  
QY 32 -----SH-----I-----S-Q-----SNI-----N-----GAL----- 43  
Db 180 EQGAPTEATAYSHLABIHSDSVILRDDFDYSQYQLELNPNIWVECSCEMGEQCGTMMHGN 239  
QY 44 -V-----P-P-----A-AL--I-----SI-----I-Q-- 54  
Db 240 AVTFCEPYGPRELTTTCLNTTTASVLQFSIGSGCRFSYSDPSITVSYAKNNTADMIQLE 299  
QY 55 -----KG-----LQY-----V-EA-E-----VSIN----- 67  
Db 300 KIRAPSNVSTVIHILYLPBEEKESVQFQWKQDSLRVGEVYACWALDNILV-INSAHRE 358  
QY 68 ---ED---GT-IF-----DG-----R----- 76  
Db 359 VVLEDNLDPDVTGNLWFFPGATVKHSCQSDGNSIYFPHGNEGSEFNATTTRDVLSTEDIQ 418  
QY 77 -----P-I-----ES-LSLI-----DA----- 86  
Db 419 EQWSEBFEQPTGWDILGAVVGADCTGVESGLSVFLKDGKRLCTPYMDTTCYGNLRFY 478  
QY 87 -VM-----P-----D-----D-----VV-----QT----- 94  
Db 479 FVWGIGCDPGVSHENDIILVAKIEGRKEHIALDTLTYSYKVPVSLVSVVINPELQTPATK 538  
QY 95 ---RQ---QAY-R---D-----K-LAQ-----Q----- 106  
Db 539 FCLROKSHOGYNNRWAVDFHVLVPLPSTMHMIQFINSINLGCCTHQPNGSVSLFSTNH 598  
QY 107 -QA-----A-----AAA-----AA----- 114  
Db 599 GRSWSLLHTECLPEICAGPHLPHSTVYSSNYSGWNRITIPLENAALTRDIRWRQTGP 658  
QY 115 -----A-----A-----A-----A-----A-----Q--GS 123  
Db 659 ILGNWAIADNVYIGPSCLKFCSGRGQCTRHGCKCDPFGSPACEMASQTFFMFISESFGS 718  
QY 124 AK-----N-----G-E-----N----- 129  
Db 719 ARLSSYHNFYSIRGAESVFCGVLASGKALVFNKGRROLITSLDSSQSRFLQTLRLG 778  
QY 130 -----T-A-----NG--E-E--N-----G- 137  
Db 779 SKSVLSTCRAPDOPGEGVLLHYSDNGITWKLLEHYVYVNYHEPRIISVELPDDARQFGI 838  
QY 138 -----A-H-----T-----I-----AN-----N-----H- 145  
Db 839 QPRWQPYHSSQGEDVWAIDEIVMTSVLFNSISLDFNLVEVTQSLGFLVGNVQPYCGHD 898  
QY 146 -----T-D-----M-----NEVDG-----D-V-E----- 156  
Db 899 WTLCTFTGDSKLASSMRVFTQSMQI-GASYMTQFSLVMGCGQKYTPHMDNQVLEYSANH 957  
QY 157 -----I-----P--EN--K-----A-----VVL-----R----- 166  
Db 958 GLTWHLVQBECLDPMFSCQEFTSASIVHASEFTQWRRTVVVLPOKTWSGATRRWQSYY 1017  
QY 167 -----GH-----E--SE-----V---F--- 173  
Db 1018 TAQDEWALDNIYIGQCCPNMCSGHGSCDHGVCRCDOGYQGTCHPEAALPSTIMSDPENP 1077



QY 174 -----I-----C-----A-----W----- 177  
Db 1078 SSWESDQWQVIGBWKVQCGVSSGSLYFSKAGKQVLVSMDLDTSVDFVQFYQI 1137  
QY 178 -----N-P-----V-----SD-----L-L-ASG----- 187  
Db 1138 GGSAAACNKPDSREBEGILLOYSNNGIQWHLAEMYPFSDKPRFVYVLELPAAGKTPCTR 1197  
QY 188 -----SG-----D-----S-----T-----A-----R-IW- 196  
Db 1198 FRWKPVSFSEDYDQWAVDDIILISEKQKQVIPPVNPPLPQNFYKPAFDYPMQMSVWL 1257  
QY 197 ---N--LSEN-S---T-S---G-S---T-----Q-L----- 209  
Db 1258 MLANEGWAKNDSFCATTPSAMVFGKSDGRFAVTRDLTLKPGVYLQKLNIGCTSQFSST 1317  
QY 210 --V-LR--H-----C-----IR-EG-----GQD-----V 222  
Db 1318 APVLLQYSHDAGMSWFLKKEGCFPASAAKCEGNSRELSPTVYTG-DPEBWTRITIAI 1376  
QY 223 P-----S-N-----K-----D----- 227  
Db 1377 PRSLASKTRFRWIOESSQKNVPPFGLDGYIISEPCPSYCSGHGDCISGVCFDLYTA 1436  
QY 228 ---V--T---S--LD---W-----NSEG-TL---L---A- 241  
Db 1437 AQGTCVSNTPNHSEMPDRPEGLKPLWKYKITGGVGTGCGTLN-DGRSLYFNLGKREAR 1495  
QY 242 T-----GS-----Y--D-G-----F 248  
Db 1496 TVPLDTRNISLVQFYIQIGTSGITVTPRAREYGLVQVSNNDGILMHLLEDFMSF 1555  
QY 249 -----A-RW-----T-----K-DG- 256  
Db 1556 LEPOIISIDLPREAKTATAFR-WWOPQHKHSAQWALGVILGVNDSQTGFQDKLDGS 1614  
QY 257 -----N-----LAS--T-L--G--Q-----H-----KG- 267  
Db 1615 IDLOANWYRLOGQVODIDL-SMDTALIFTEIGNRYAETWDFHVSSESLQWENMGC 1673  
QY 268 --PIP-A--L-----K-W--NKK-----G----- 278  
Db 1674 SKP-FSGAHGIQLOYSINNGKQWLTVECPPTIGCVHYTESSTYTSERFQWRRVTV 1732  
QY 279 -----N-----F--I--LSAGYDKT-TI--I-----W-----DA----- 295  
Db 1733 LPLATNSPRTRFRWIOQNTYTVGAD-SWAIDNVILASGCPMWCGRGICSDRGVCDRGFG 1791  
QY 296 -----H-----T-----GE----- 299  
Db 1792 GPFCVPVPLPSILKDDPFGNLHPDLWPVYVGAERGNLNGETIKSGTCLIFKGEGRMLI 1851  
QY 300 -----AK--Q--QF-----P-----H-----S----- 308  
Db 1852 SRDLCTNTWYQVSLRFIAKGTPEKSHSILLQFSVSGVTWHLMDDEFYPPQTTSLFIN 1911  
QY 309 AP-----A--L-----D-----V-----D-W 316  
Db 1912 VPLPYGAQTNATFRMLQWPNYNGKBEIWIIDDFIDGNLNNPVLILLDTDFGPREDNW 1971  
QY 317 -----Q--S-N-NT----- 321  
Db 1972 FFYPGGNIGLYCPSYSGAPEDSAMVFSVNEGEHSITTRDLSVNENTIIQPEINVGCS 2031  
QY 322 -----F-A-----S-CST----- 327  
Db 2032 TDSSADPVRLEPFRDGFATWHLPLCYHSSSLVSLCSTGHEHPSSTVYAGTTQWRRRE 2091  
QY 328 -----D-----MC-IH-----VC 334  
Db 2092 VWHFGKLHLCGSVRFRWYQFYPAGSQPVTWAIIDNVYIGQCCEMICYGHGSCINGTKCIC 2151

QY 335 -----K--L---L---GO---DR-----P-----I-KT-----P-Q-G- 347  
Db 2152 DPGYSGPTCKISTKNPDFLKDDEGQLESDFLLMSGKPSRKCGILSSGNLFFNEDGL 2211  
QY 348 -----H-----T-N-----EV----- 352  
Db 2212 RMLVTRDLDLSHARFVQFMRLGCGKGVPPRSPQVLLQYSLNGLSWSLQLEFLFSNSS 2271  
QY 353 NA--I--K-----W----- 357  
Db 2272 NVGRYIALEMPKARSSTRLRWQPSENGHFYSPWVIDOILIGGNISGNTVLEDDFSTL 2331  
QY 358 D-----P---TG-----N---LL-----ASCS--D 369  
Db 2332 DSRKLLHPGGRMPVCGSTGDALVFEKASTRYVVTDDIADVNEDSLQIDFAASCSTVD 2391  
QY 370 -----DM-----T-----L-KI-----W----- 376  
Db 2392 SCVAIELEYSVDLGLSMHPLVRDCLPTNVECSRYHLQRLILVSTFNKWTTRITLPLPSYTR 2451  
QY 377 S-----M-----KO---DN---CV---H-----D--- 386  
Db 2452 SQATRRWHQAPAPFDKQOTWAIIDNVYIGDCLDMCSGHGRCVCGSCVDEQWGLYCDPE 2511  
QY 387 ---L--Q-----Q-----H----- 390  
Db 2512 ETSPLTQLKDNFRAPSNQNLTVSGKLSLTVCGAVASGLALHFGSGCSRLLVTDVNLNT 2571  
QY 391 NKE-I---Y---T-----IKW-----S-PTG-----P 404  
Db 2572 NAEFIOFYMYGCLITPFSNRNOGLLEYSVNGGITWLLMEIFYDQYSKP-GEVNTILLPP 2630  
QY 405 ---CT-----NNP-----N--A--NMLA-SA-----SF-----D 421  
Db 2631 DAKBIATRRWQ--PRHGDLDQNDWAIIDNVILSGSADQRTVMDLTSSAPVPOHERSPAD 2689  
QY 422 S-TV-R-----L-----W---DVD---R-----GI--C-IH---TL--TKH- 441  
Db 2690 AGPVGRIAFEMFLEDKTSVNEENWLFHD-DCTVERPCDSPDGMVLCGSHDREVAVT-HD 2747  
QY 442 ---Q-----E-P-----V-YSAF-----S---P-D----- 453  
Db 2748 LPTENWIMQFKISVGCKVPEKIAQNIHVQFSTDFGVSWLYVPOCLPADPCKSGSVSQ 2807  
QY 454 -----G--R--Y-L-AS--G-----S-----F-----D- 463  
Db 2808 PSVFPFTEGKRITYPLESITGNVFRFPYQKYSVQWAIIDNVYIGPCGLDNCGGHGDC 2867  
QY 464 -K--CV-----H-----I---W-----NT- 471  
Db 2868 LKEQCICDPGSGPNCYLTHSLKTLKRFDSSEEIKPDLWMSLEGGSTCTCECVLAENTA 2927  
QY 472 ---Q--T-----GA--L---V-----H-----S----- 479  
Db 2928 LYFGSTVQAQITQDILDRGAKFLOWGRIGSENNTSCHRVPCKEGLVLLDFSTGGIT 2987  
QY 480 -----Y--R-----G-----T----- 483  
Db 2988 WTLHHEMDFQKYSVVRHDYILLPEGALTNTTRLRWQPFVINSGLVSVGVERAQWALDNI 3047  
QY 484 --GG--I-----F-----E-----V---C-----W-----NA 493  
Db 3048 LIGGAENPSQLVDFDDESGSSHEENWSFYPNAVRTAGFCGNPSFHYLPNPKKDKTHNA 3107  
QY 494 -----A---GD-----K----- 497  
Db 3108 LSSRELIQGYMOPKIVVGCEATSCGLHSVMLEYTKDARSWSQLVQTCPLPSSNS 3167  
QY 498 VG-----A-----SA-----S---D--- 504  
Db 3168 ICGSPQFHEATYNAVNSSWKRTIQLPDHVSSATOFRWIQKEETEKOQSWAIDHVY 3227  
QY 505 -G-----S-----VC----- 508



Db 1140 EKNCNSTETCOPSFQFNCNPHRCHIDLSEVCDGDKDVCVDSDEVGCVLNCNTASQFKCASGDK 1199  
Qy 163 ---V---VL---R-G-H---E--- 169  
Db 1200 CIGVTRCDGVFDCSDNSDEAGCPTPPGCHSDEBFCQBDGICIFNFWCEDGHPDCLYG 1259  
Qy 170 S-E---V---P---IC---AW---N---P--- 179  
Db 1260 SDEHNACVPKTCPSYFHCDCNGCIIHRAWLDCDRNDCGMSDEKDCPTQPFRCFSQWQC 1319  
Qy 180 ---V---V---SD--- 182  
Db 1320 LGHNICVNLVCDGIFDCPNCGTDESPLCNGSCDFNGCGTCHECQVEPFGAKCLPLGF 1379  
Qy 183 LLA---S---GS---G--- 189  
Db 1380 LLANDSKTCEDIDECILGSCSHCYNMRGSPRSCDTGYMLESDBRTCKVTASESLLLL 1439  
Qy 190 ---DS-T---A---RI---W---N--- 197  
Db 1440 VASONKIIADSVTSQVHNIYSLVNGSYIIVADPDSISGRIFWSDATQGTWAFQNGTD 1499  
Qy 198 ---LSE---N---S---T---S---GS---T---Q-LVL 211  
Db 1500 RRVVFDSSIILTETIADWVGRNLYWTDYALETIEVSKIDGSHRTVLISKULTNPRGLAL 1559  
Qy 212 ---R---H---Q--- 213  
Db 1560 DPRNEHLLFWSDWGHHPRIERASMDGSMRTVIQDKIFWPCGLTIDYPNRLLYFMDSYL 1619  
Qy 214 ---C---IR---E---GG--- 219  
Db 1620 DYMFCDYNGHRRQVIASDLIIRHPVALTLFDSVYVTTDRATRVMRANKWGGQSVV 1679  
Qy 220 ---Q--- 220  
Db 1680 MYNIQWPLGIVAVHPSKQPNVNCFAFRCRSHLCLLSQGPFPYFCVCPGSGWSLSPDLLN 1739  
Qy 221 ---D---V---P---SN-K---DV--- 228  
Db 1740 CLRDDQFLITVQHOIIFIGISLNPEVKSNDAMVPIAGIQLGLDVEFDDAEQYIYWVENPG 1799  
Qy 229 ---T---S---LDW---N---S-E---G---T 238  
Db 1800 EIHVRKTDGNTRVFASISVMGFSMNLALDWISRNLYSTNPRTQSIEVLTLHGDIYRKT 1859  
Qy 239 LLA---T-G---S---Y---DG---F--- 248  
Db 1860 LIANDGTALGVGFPFIGITVDPARGLYWSQDGTDSGVPAKIASANMDGTSVKTLFTGNLE 1919  
Qy 249 ---A---R-I---W---T--- 253  
Db 1920 HLECVTLIDIEEQKLYWAVTGRGVIERGNVDGTRMILVHQLSHPWGIAVHDSFLYYTDEQ 1979  
Qy 254 ---K-D---G-N---L---ASTL-G-Q-H---K--- 266  
Db 1980 YEVIERVDAKGANKIVLRDNVENLAGLQVYHRRNAESSNGCSNNMACQQICLPVPG 2039  
Qy 267 ---G---P---I---FALK--- 273  
Db 2040 LFSCACATGKLPNDRSCSPYNSFIWVSMLSAIRGFSLELSDHSETMVPVAGQGRNALH 2099  
Qy 274 ---W---N---K---K---GN- 279  
Db 2100 VDVVSSGFIYWCDFSSVASDNAIRRIKPDGSSLMNIIVTHGIGENVRGIAVDWVAGNL 2159  
Qy 280 ---F---I--- 281  
Db 2160 YFTNAFVSEFLIEVLINTYRVLKVTVMPRHIVVDPKNRYLFWADYQGPPIERSF 2219  
Qy 282 ---L-S---A---G---VD--- 287  
Db 2220 LDCTNRTVLVSEGIIVTPRGLAVDRSDGYVYVWVDDSLDIIRINGENSEVIRGSRYP 2279

Qy 288 ---K---T---T-I---I-W---D--- 294  
Db 2280 PYGITYFENSIIWDRNLKKIPOASKEPENTTEPTVIRDNINWLRDVTIFDKQVQPRSPA 2339  
Qy 295 ---A---HT---G---E---A--- 300  
Db 2340 EVNNPCLENNGCGSHLCPALPGLHTPKDCAPFTLQSDGKNCAISTENFLIFALSNSLR 2399  
Qy 301 ---K---Q---Q---F--- 305  
Db 2400 SLHLDPENHSPPPQTINVERTVMSLDYDSVSDRIYFTQNLASGVQISYATLSGIHTPT 2459  
Qy 306 ---F--- 306  
Db 2460 VIASGIGTAGIAFDWITRIIYSDYLNQMINSMAGDSNRTVIARVPKPRAIVLDPGCG 2519  
Qy 307 ---H---S-AP-ALD- 313  
Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSSLVMPGSLTLDYEEDLLYWDASLQRIERST 2579  
Qy 314 ---VD---W---Q---S---N--- 319  
Db 2580 LTGVDRVIVNAVAHAFGLTYQYIYTDLYTQRIYRANKYDGSQGOIAMTTLNLSQPRG 2639  
Qy 320 ---NT---F--- 322  
Db 2640 INTVKNQKQCCNPNCEQFNGGCSHICAPNGABEQCPHEGWNWYLANNRKHCIVDNGER 2699  
Qy 323 ---AS---CS---TD-M---C-IH---V--- 333  
Db 2700 CGASSPTCSNGRCISBEWKCDNDGCGDSESVICALHTCSPTAFTCANGRCVQVSYR 2759  
Qy 334 CKL---G-OD---R-P-I--- 342  
Db 2760 CDYNDGCGSDEAGCLFRDCNATTEFCMNRRCIPREFICNGVDNCHDNTSDEKNCPD 2819  
Qy 343 ---K---K---T---T---FO--- 346  
Db 2820 RYCQSGYTKCHNSNICIPRYLDCGNDGCDNSDENPTYCTHTCSSEFQASGRICIPQ 2879  
Qy 347 ---GH---T--- 349  
Db 2880 HWYCDQETDCFDASDEBPASCGHSERTCLADEFKDCGGRICIPSEWICDGDNDGCDMSDEK 2939  
Qy 350 ---N---E---VNA---IK-W--- 357  
Db 2940 RHQCQNQCSDEFLCVDNDRPPDRRCIPOSWVCDGSDVDCDGDYDENQNCNTRTCSENEFT 2999  
Qy 358 ---LAS-CS---D-D-M--- 357  
Db 3000 CGYGLCIPKIFRCDRHNDGCDYSDERGCILYQTCQOQFTCQNGRCISKTFVCDENDDCGD 3059  
Qy 358 ---D---PT---G---N---L--- 363  
Db 3060 GSDELMHLCHTPEPTCPPEFKCDNGRCIEMMKLCNHLDDCLDNDSEKKGCGINECHDPSI 3119  
Qy 364 ---LAS-CS---D-D-M--- 371  
Db 3120 SGCNDHNTDTLTSFYCSCRPGYKLMDSKRTCTVDIDECTEMPFCVCSQKCNENVISYICKCA 3179  
Qy 372 ---T---L--- 374  
Db 3180 PGLREPKDCTCRONSIEPYLIFSNRYLYRLNLTIDGYFYSLLILEGLDNVVALDFORVEK 3239  
Qy 375 ---I---W---S--- 377  
Db 3240 RLYWIDTQRIERFPLNKTNETIINHRLPAAESLAVDWVSRKLYWLDAERLDFVSDL 3299  
Qy 378 ---M-KQ---D-N-C---VHDLOO---H---NK- 392  
Db 3300 NGCHRRMLAQHCVDANNTFCFONPRGLALH---POGYLYWADWGHGRAYIGRVGMDGTNKS 3357

QY 393 -----E-----I-YT-----IK----- 398  
Db 3358 VIISTKLEWNGITIDYTNLLYWAHAHLGYIEYSDLEGHHRHTVYDGLPHFAITIFE 3417  
QY 399 -----W-----S-----P-T--G-- 403  
Db 3418 DTIWTWNTTRTEKNGKDGSRQTLVNTTHRPFDIHVHYPRQPIVSNPCGTNGGCS 3477  
QY 404 -----PG-----T-----NN----- 408  
Db 3478 HLCILKPGKGFTCECPDPDFRTLLQSGSTYCMPCMSSTQFLCANNKRCIPIMWKCDQKD 3537  
QY 409 -----P-----NA-N-----LML-----AS 417  
Db 3538 CSDGSDRLALCPQFBLGQFQSGDNGCTSPQTLNAHQNCPCGSDBDRL-LCENHHCD 3596  
QY 418 -----A-----S-----F-----DS-----T-----VR-----L-----W-DV 429  
Db 3597 NEWQANKRCIPESWQCDTFNDCEDESDSDSHCASRTCPGQFRANGRCIPQAWKCDV 3656  
QY 430 --DRG-----I-CI--H-----T-L----- 438  
Db 3657 DNDGDSHSDPIECMSSAHLCDNFTEFSCKTYRCIPKWAVCNVGDVDCRDNDSDEQCEE 3716  
QY 439 -T-----K-H-----Q-----E-P-----V----- 445  
Db 3717 RTCHPVGDPRCKNHCIPLRWQCGQNDGSDGNDSENCAPRECTESEFRVCVNOQCTPSRW 3776  
QY 446 -----Y-----S-----V-----A-----F 450  
Db 3777 ICDHYNDGSDNDBERDCMRTCHPEYFQCTSGHCVSHSELKCDGSADCLDASDEADCFTRF 3836  
QY 451 SPDGYL-AS-----GS-----P----- 462  
Db 3837 -PDGAYCOATMFECKNHVCIPPYWKCDGDDCGDSDDEHLCLDVFCSNPFRCNDNR 3895  
QY 463 -----D-----K-----CV-H--I----- 468  
Db 3896 CIYSHEVCNVGDDCGDGTDETEHCKRPTKPTCYEYKCGNGHCIPHNDVNCDDADDGCD 3955  
QY 469 W-----N-----T-----Q-T-----G-----A-----L-V----- 477  
Db 3956 WSDGLGCKGKERTCAENICQNTQNLNEGFTCSCTAGFTNVFDRTSCLDINECEBQFG 4015  
QY 478 -----H-----SY-----R----- 481  
Db 4016 TCPQHCRNTKGYECVACDGTSMDSRPGKRCAAEGSSPLLLLPDNDVNRKYNLSSEF 4075  
QY 482 -----G-----T-----G-----G-----I-FE-----V-C-- 490  
Db 4076 EYLDEEYIQAVDYDWDPKDIGLSVYVYTVRGESRFGAIAKRAYIPNFESGRNLLVQEV 4135  
QY 491 -----W-----N-----AA----- 494  
Db 4136 LKLKVMQPDGIAVDWVGRHRYNSDVKNKRIEVAKLDRYRKWLITDLDQPAIAVNP 4195  
QY 495 -----G-----G-----D-K-V----- 498  
Db 4196 LGLMFTWDGKPKIESAWNGSDRNILVFDLGMFTGLSIDYLNDRYWSDFKEDIET 4255  
QY 499 -----G-----A-SA-----S-----D-----G----- 505  
Db 4256 IKYDGTDRRVIKAEAMNPYSIDIFEDQLYWISKEGEVWKNKFGQKKEKTLVVPNPLT 4315  
QY 506 -----SV-----C-V-----L-----D----- 511  
Db 4316 QVRIFHOLRYNKSVPNLCKQICSHLCILRLPGYSCACPOGGSFIEGSTRTECDAAIELPIN 4375  
QY 512 L-----R-----K 514  
Db 4376 LPPFCRCMHGNCYFDETDLPK 4397

RESULT 5  
US-08-652-877-84  
; Sequence 84, Application US/08652877  
; Patent No. 6187548  
; GENERAL INFORMATION:  
; APPLICANT: Akerstrom, Goran  
; APPLICANT: Jublin, Claes  
; APPLICANT: Rask, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjaln, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
; TITLE OF INVENTION: Thereof and DNA Encoding Same  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd., 3C43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.5.1  
; SOFTWARE: Word 6.0 (Patentin)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/652,877  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: PCT/US95/15203  
; APPLICATION NUMBER: PCT/US95/15203  
; FILING DATE: 22-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/344,836  
; FILING DATE: 23-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,314  
; FILING DATE: 07-JUNE-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitzky, Martin  
; REGISTRATION NUMBER: 29,699  
; REFERENCE/DOCKET NUMBER: A1355E-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-454-3816  
; TELEFAX: 610-454-3808  
; INFORMATION FOR SEQ ID NO: 84:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4655 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-652-877-84

Query Match 70.1%; Score 2566.6; DB 3; Length 4655;  
Best Local Similarity 9.8%; Pred. No. 5.1e-32;  
Matches 433; Conservative 57; Mismatches 19; Indels 3894; Gaps 381;

QY 1 M-----S-----I-----S--SDE- 7  
Db 1 MDRGPAVACTLLALVACLAPASGQECDSAHPRGSGHCIPADWRCDGTGKDCSDDADEI 60  
QY 8 ----VN-----F-----L-----V----- 12  
Db 61 GCAVWTCQGYFKQSEGCIPSSWYCDQDQDCDSDRQDCSQSTCSHQITCSNGQC 120  
QY 13 ----YR-----Y-----L-----Q-----E--S 19  
Db 121 IPSEYKCDHVRDCPDGADENDCOYPTCEQLTCNGACYNYSOKCDWKVCDRSDSDEINCT 180  
QY 20 ----G-FS-----H-----S--A--F-T-----G-I----- 30

Db 181 EICLHNEFCGNGECIPRAYVVDHNDQDGSDEHACNYPTCGYQYTCPSGRCIYQNWV 240  
QY 31 K-----S-----H-----IS-----O-SNI 39  
Db 241 CDGEDDKDNGDEDCESGPHDVHKCSPREWSCPSGRCISIVKVCDDIILDCRGREDN- 299  
QY 40 N---G---AL---VP---PAALI---S---I-I-- 53  
Db 300 NTSTGKYSMTLCSALNCQYCHETPYGGACFPYIINHNSRTCVDFDQOIMGICD 359  
QY 54 QK-----G--L--QY-----VE 61  
Db 360 QKCESRPGRLHCHCEGYILEROYQYKANDSFGASIFNSGRDLLGDIHGRSFRILVE 419  
QY 62 -----A-----E---VSI-----N 67  
Db 420 SQNRGAVGVAFYHHLQRFVTDVQNVKVSVDINGLNIQEVNLVSVETPENLAVDWVN 479  
QY 68 -----E-----DG---TL-----72  
Db 480 KIVLVETKVRIDWNLDSYRVTLITENLGHPRGIAVDPTVGYLPFSDWESLSGEPKLE 539  
QY 73 --F-DG--R-----P-----IE-----S 80  
Db 540 RAFMDGSNRKDLVKTLGWPAGVTLDMISKRVYVWDSRFDYIETVYDGIQRKTVVHGS 599  
QY 81 L-----SL--I--D---AVM-----88  
Db 600 LIHPFGVSLFEGOVFTDTKMAVLKANKFTETNPQVYQASLRPYGVTVYHSLRQPYA 659  
QY 89 --P--D-----V--V-----Q--T--R--Q-----Q-AY 99  
Db 660 TNPCKDNNGCEQCVLSHRTDNDGLGFRCKCTFGFOLDTDERHACIAVQNLFPSSQVAI 719  
QY 100 R-----D-K-----L-AQ-----Q-----Q--A 108  
Db 720 RGIPFTLSTQEDVMVPVSGNPSFFVGIDFDAQDSTIFFSDMSKHMIFKQIDGTGREILA 779  
QY 109 A-----A-----A-----A-----A-- 112  
Db 780 ANRVENVESLAFDWISKVLTWDSHYKSI SVMLADKTRRTVVOYLNPNRSVVHPFAGY 839  
QY 113 -----A-----A-----AA-----S-- 119  
Db 840 LFPTDFRPAPKIMRANGSGSHLLPVNTILGPNGLAIDWASRLYWDVAFDKIEHSTF 899  
QY 120 -----Q-----120  
Db 900 DGLDRRLGHIEQWTHPFGLAIFGEHLFFTDWRLGAIIRVRKADGEMTVIRSGIAYILH 959  
QY 121 -----Q--S-A-----KNG-----EN-----TA-N-----G-----133  
Db 960 LKSYDNIQTSNACNPTHPNGDCHSFPPVFNFORVCGPYGMRLASNHLTCGDPPTN 1019  
QY 134 E---E---N-----N-----GA--H-----T-----140  
Db 1020 EPTEQGLFSFPCKNGRCVPNYLDCGVDDCHNSDEQLCGTILNNTCSSAFTCGHGEC 1079  
QY 141 I-A-----N-----N-----H-----T--DME-----V-D--GD-- 154  
Db 1080 IPAHWRCDKNDKVDGSDHNCPTHPASCLDTQYTCNHCQISKNWVCDTNDGCGSD 1139  
QY 155 -----VEI--P-----PN-----R-G-H-----E-----K-A-----162  
Db 1140 EKNCSNETQPSQFNCNPHRCIDLSPVCDGDKDVGSDVGVCLNCTASQPKCASGDK 1199  
QY 163 -----V-----VL-----R-----G-----E-----169  
Db 1200 CIGVTRCDGVFDCSDNSDEAGCPTRPFGMCHSDEFQCCEDGICIFNFWECGHPDCLYG 1259  
QY 170 S-E-----V-----F-----IC--AW-----N-----P-----179  
Db 1260 SDEHNACVPKTPSSYFHCNNGNCIHRALICDRDNDGDMDSDEKDCPTQPFRCPSQWQC 1319

QY 180 -----V-----SD-----182  
Db 1320 LGHNICVNLVVDGIFDCPNGTDESPLCNGNSCSDENGCGTCEVCQEPFGAKCLPLGF 1379  
QY 183 LLA-----S-----GS-----G-----189  
Db 1380 LLANDSKTCEDIDECIDILGSCSQHCYNMRGSRFCSDTYMLSDSGRTCKVTASESLLL 1439  
QY 190 -----DS--T-----A-----RI-----W--N-- 197  
Db 1440 VASQNKIIADSVTSQVHNIYSLVNGSYIVAVDFDSISGRIFWSDATQGTWSAFNGTD 1499  
QY 198 -----LSE-----N--S--T-----S--GS-----T--Q--LVL 211  
Db 1500 RRVVFDSSILTTETIAIDWGRNLYTWDYALETIEVSKIDGSHRTVLISKNLNPRLGAL 1559  
QY 212 --R-----H-----Q-----213  
Db 1560 DPRMNEHLLFWSDWGHHPRIERASMDGSMRTIVQDKIFWPCGLTIDYPNRLLYFMDSYL 1619  
QY 214 -----C-----IR-----E-----GG-----219  
Db 1620 DYMDFCDYNGHRRQVIASDLIIRHPYALTFLFEDSVYWTDRTRVRMRANKMHGQSVV 1679  
QY 220 -----Q-----220  
Db 1680 MYNIQWPLGIVAVHPKQPNVNPFCASRCSHCLILSSQGPFPYSCVPSGWSLSPDLLN 1739  
QY 221 -----D-----V-----P--SN--K-----DV-----228  
Db 1740 CLRDDQPFLLITVRQHIIFGISLNPVKSDAMVPIAGIQLGDLVEFDDAQIYVWENPG 1799  
QY 229 -----T-----S-----LDW-----N--S--E--G-----T 238  
Db 1800 EHRVKTGNTRTVFASISWVGSMNLALDWSIRNLSTYNTQSIETVLTHGDIRYKT 1859  
QY 239 LLA--T--G-----S-----Y-----DG-----F-----248  
Db 1860 LIANDGTALGVGPPIGITVDPPARKLYWSDQGTGVPKIASANMDGTSVKTLFTGNLE 1919  
QY 249 -----A-----A-----R-I-----W-----T-----253  
Db 1920 HLECVTLDIEOKLYWAVTGRGVIERGNDGTDMLVHQLSHPWGIAVHDSFLYTTDEQ 1979  
QY 254 --K--D--G--N--L--ASTL--G--Q--H-----K-----266  
Db 1980 YEIVERVKATGANKIVLRDNDVNLRLQVYTHRRNAESSNGCSNNMNACQICLPVPGG 2039  
QY 267 -----G-----P-----P-----I--FALK-----273  
Db 2040 LFSACATGKLPNDRSCSPYNSFIVMSLSAIRGFSLELSDHSETWVPVAGQGRNALH 2099  
QY 274 -----W-----N-----K-----K-----GN--279  
Db 2100 VDVVSSGFIYWCDFSSVASDNAIRRIKPDGSSLANIVTHGIGENGVRGVIADWVAGNL 2159  
QY 280 -----L-S-----A-----G--VD-----F-----I-----281  
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTVDMPRHIVVDPPKNRYLFWADYQORPKIERSF 2219  
QY 282 -----L-S-----A-----G--VD-----287  
Db 2220 LDCNRTVLVSEGIPTVPRGLAVDRSDGYVYVWDDSLDIATIRINGENSEVIRYGRYPT 2279  
QY 288 -----K-----T-----T--I--I--W-----D-----294  
Db 2280 PYGITVENSIIWDRNLKKIQAOSKEPENTPEPTVIRDNINWLVDVTFDKQVQPRSPA 2339  
QY 295 -----A-----A-----HT-----G-----E-----A-----300  
Db 2340 EVNNPNPLENNGGCHLCPALPGLHTPKDCAPGTLQSDKGNCAISTENFLIPALSLSLR 2399

301 QY -----K-----Q-----Q-F-----P-305  
2400 Db SLHLDENHSPPTQINVERTVMSLDYSDRIYFTQNLASGVGQISVATLSGHIHTPT 2459  
306 QY -----F----- 306  
2460 Db VIASIGTADGIAFWITRIIYSDYLNQMINSGMAEDGSNRTVIARVPRAIPLVDPQCG 2519  
307 QY -----H-----S--AP--ALD----- 313  
2520 Db YLYWADWTHAKIERATLGNFRVPIVNSLVMPGSLTLDYEBLLYVWDASLQRIERST 2579  
314 QY -----VD-----W-----Q-----S--N----- 319  
2580 Db LTGVREVIYVAAHAFGLTYGYIYWTOLYQRIYRANKYDGSQGIAMTTLNLSQPRG 2639  
320 QY -----NT-----F----- 322  
2640 Db INTVWKQKQCNPCQFNGGCSHCAPGPGAECQCPHEGNWYANNRKHCIVDNGER 2699  
323 QY -----AS-----CS-----TD-M--C-IH-----V----- 333  
2700 Db CGASSFTCSNGRCISEWKCNDNDGDSDEMSVCAHTCPTAFTCANGRCVQYSYR 2759  
334 QY -----G--QD-----R-P--I----- 342  
2760 Db CDYNDGCGSDGAGCLFRONATTEPMCMNRRCIPREFICNGVNDCHDNTSDEKNCPS 2819  
343 QY -----K-----T-----FQ----- 346  
2820 Db RTCQSGYTKCHNSNICIPRVVLCDGNDGNDSDENPTYCTTHTCSSSEFQACSGRCIPQ 2879  
347 QY -----GH--T----- 349  
2880 Db HWYCDQETCFDASDEPASCBSHRTCLADEPKDGRGRCIPSEWICDGDNDGDSDEK 2939  
350 QY -----N--E--VNA-----IK--W----- 357  
2940 Db RHQONQNSDSEFLCVNDRPDRRCIPQSVWCDGVDCDGDYENQNCRTTCSENEFT 2999  
358 QY ----- 357  
3000 Db CGYGLCIPKIFCRDHNDCGYSDERGCLYQTCQOQFTCQNGRCISKTEVCDEDNDGCD 3059  
358 QY -----D-----PT-----G-----N-----L----- 363  
3060 Db GSELMHLCHTPEPTCPHFKDNGRCIEMMKLCNHLDCDLDNSDEKGCINECHDPSI 3119  
364 QY -----LAS--CS-----D-D--M----- 371  
3120 Db SGCDHNCITLTSFYCSCRPGYKLSMDKRTCVDIDECTEMPFVCSQKCVNIGSYICKCA 3179  
372 QY -----T-----L-----K 374  
3180 Db PGLYREFDGTQRONSNIPIYFSNRYLNLRTIDGYFYSLLGLELDNVVALDFDREK 3239  
375 QY -----I-----W-----S-- 377  
3240 Db RLYWIDTQRVIERMFLNKTNETIINHLPAESLAVDWVSRKLYWLDARLDGLFVSD 3299  
378 QY -----M-KO--D--N--C-----VHDLQO-----H-----NK- 392  
3300 Db NGHRRMLAQHCVDANTTFCDNPRGLAH--PQYGLYWADWCHRAYIGRVGMDGTNKS 3357  
393 QY -----E-----I-YT-----IK----- 398  
3358 Db VIISTKLEWPNGITIDVTLNLLYWADAHGLYIEYSDLEHGHRTVYDGPALPHPAITIFE 3417  
399 QY -----W-----S-----P--T--G-- 403  
3418 Db DTIYWDWNTRTVEKNGKDGNSRQTLVNTTHRPFDIHVYHPYQPIVSNPCGTFNNGCS 3477  
404 QY -----PG-----T-----NN----- 408

3478 Db HLCLKPGKGFTCECPDDFRTLQLSGSTYCMPCSSSTOFLCANNEKCIPIWMKCDGQKD 3537  
409 QY -----P--NA--N-----LML-----AS 417  
3538 Db CSDGSEALCALCPQRCRLGQFQCSGNGNCTSPOTLCNAHONCPDGSDEDL-LCENHHDS 3596  
418 QY -----A-----S-----F-----DS-----T-----VR-----L--W--DV 429  
3597 Db NEWQCANKECIESWCQDTFDCEDNSDESHCSARTCPGQFCRANGRCIPOAMKCDV 3656  
430 QY -----DRG-----I--CI--H-----T-L----- 438  
3657 Db DNDGDSDEPIEECMSSAHLCDNPTFESCKTNYRCIPKWAVCNGVDDCRDSDGQCEE 3716  
439 QY -----T-----K--H-----Q-----E--P-----V----- 445  
3717 Db RTCHPVGDPRCKNHHCIPLRWQCDGQNDGNDSDENECAPRECTESEFRVCVNOQCIPSRW 3776  
446 QY ----- 450  
3777 Db ICDHYNDGDSRDERDCENRTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCPTRF 3836  
451 QY -----SPDGRYL-AS-----GS-----F----- 462  
3837 Db PDGAYCATMECKNHVCIPPYWKCDGDDCGDSDEBLHLCLDVPCNSPNRFRCDNNR 3895  
463 QY -----D-----K-----CV--H--I----- 468  
3896 Db CIYSHEVNGVDDCGDGTDETEHCKRTPKCTEYECGNGHCIPHONVDDADDGCD 3955  
469 QY -----W-----N-----T-----Q--T-----G-----A-----L--V----- 477  
3956 Db WDELGNCKGKERTCAENICEQNCITQNEGFTCSCTAGETNVFDRTSCLDINECEQFG 4015  
478 QY -----H-----SY-----G-----T-----G--G--G--I--FE-----V--C-- 481  
4016 Db TCPQHCNRTKGSYECVCADGFTSMDSRPGKCAAGSSPLLILPDNVIRKYNLSERFS 4075  
482 QY ----- 490  
4076 Db EYLQDEYIQAVDYDWDPKDIGLSVVYTVRGESRFGAIKRAYIENFESGRNVLQOEVD 4135  
491 QY -----W-----N-----AA----- 494  
4136 Db LKLYVMQPDGLAVDWGRHYIWSDVKNKRIEVALKLDGRYKWLISTDLDPAAIAVNP 4195  
495 QY -----G-----A-SA--S-----D-----D-K--V-- 498  
4196 Db LGLMFWTDWKEPKIBSAWMNGEDRNILVFDLWPTGLSIDYLNNDRIYWSDFKEDVIE 4255  
499 QY -----G-----A-SA--S-----D-----G----- 505  
4256 Db TIKYDGTDRRVIKAEAMNPYSLDIFEDQLYWISKEKEVWKNQKFGGKKEKTLVNPWL 4315  
506 QY -----SV--C--V-----L-----D----- 511  
4316 Db TVRIFHOLRYNKSVPNLKQICSHLCLEPPGYSACPOGSSFIESTTECDAAIELPI 4375  
512 QY -----L-----R-----K 514  
4376 Db NLPPCKRMHGGNCYFDETDLPK 4398

## RESULT 6

US-08-652-877-88  
; Sequence 88, Application US/08652877  
; Patent No. 6187548  
; GENERAL INFORMATION:  
; APPLICANT: Akerstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Raak, Lars  
; APPLICANT: Crumley, Gregg R.



QY 190 ---DS-T---A---RI---W---N--- 197  
Db 1440 VASONKIIADSVTSQVHNIYSLVNGSYIVAVDFDSISGRIFWSDATQGTWAFONGTD 1499  
QY 198 ---LSE---N---S---T---S---GS---T---Q---LVL 211  
Db 1500 RRVVDFSSIIITETIAIDWGRNLYWTDYALEITEIVSKIDGSHRTVLISKNLNPRLAL 1559  
QY 212 -R-----H----- 213  
Db 1560 DPMNEHLLFWSDGHHPIERASMDGSMRTIVQDKIFWPCGLTIDYPNRLLYFMDSYL 1619  
QY 214 ---C---IR---E--- 219  
Db 1620 DYMDPCDYNHRRROVITASLIIRHPYALTFLFDSVYVTRATRRVRANKWEGNQSVV 1679  
QY 220 -----Q----- 220  
Db 1680 MYNIQWPLGIVAVHPKQPNVPCAFSRCSHLCLLSQGPHFYSCVPCSGWSLSPDLLN 1739  
QY 221 ---D---V---P---SN---K---DV----- 228  
Db 1740 CLRDDQPLITVRQHIIFGISLNPVKNSDAMVPIAGIONGLDVEFDDAEQIYIYVENPG 1799  
QY 229 ---T---S---LDM---N---S---E---G---T 238  
Db 1800 EIHRTVDGNRTVFASISVMGSPSMNALOWISRNLYSTNPTQCSIEVLTLHGDIYRKT 1859  
QY 239 LLA---T---G---S---Y---DG---F--- 248  
Db 1860 LIANDGTALGVGPIGITVDPARGKLYWSDQDTSVGPAKIASANMDGTSVKTLFTGNLE 1919  
QY 249 -----A-----R-I---W-----T--- 253  
Db 1920 HLECVTLDIEQKLYMAVTCRGVIERGNVDGTRMILVHOLSHPWGIAVHDSFLYTTDEQ 1979  
QY 254 ---K---D---G---N---L---ASTL---Q---O---H---K--- 266  
Db 1980 YEIVERTKATGANKIVLRDNVPLRGLQVYHRRNAEISSNGCSNNWACQIICLPVPGG 2039  
QY 267 ---G---P-----I---FALK----- 273  
Db 2040 LFSCACATGKLPNDNSCSPYNSFIVVSMLSAIRGFSLESDHSETMVPVAGQGRNALH 2099  
QY 274 ---W-----N---K-----K-----GN- 279  
Db 2100 VDVVSSGFTYWCDFSSSVASDNAIRRIKPDGSSLMNIVTHGIGENGVRGIAVDWVAGNL 2159  
QY 280 -----F-----I----- 281  
Db 2160 YFTNAFVSELTIEVLRINTYRRVLLKVTVDMPRHIVVDPKNRYLFWADYQRPKIERSF 2219  
QY 282 ---L---S---A---G---VD----- 287  
Db 2220 LDCTNRTVLVSEGIPTVPRGLVADRSQVYVWDDSLIARIINGENSEVIRYGRYPT 2279  
QY 288 ---K---T---T---I---I---W-----D----- 294  
Db 2280 PYGITVFENSIWVDRNLKXIFQASKEPENTPEPTVIRDNINLMDVTIIFDKQVQPRSPA 2339  
QY 295 ---A---HT-----G-----E---A--- 300  
Db 2340 EVNNPCLNNGGSHLCFALPGLHTPKCDCAFGLTQSDGKNCAISTENFLIFALNSLR 2399  
QY 301 ---K---Q-----Q---F-----P- 305  
Db 2400 SLHLDPNHSPFQTIINVERTVMSLDYSDVSDRIYFTQNLASGVQIISYATLSSGIHTPT 2459  
QY 306 -----F----- 306  
Db 2460 VIASGIGTADGIAFDWTRRIYYSYDNLNQMINMAEDGSNRTVIARVPKPRAILVLDPCQG 2519  
QY 307 -----H-----S---AF---ALD----- 313

Db 2520 YLYWADWDTHAKIERATLGCGRFVRPIVNSSLVMPSSGLTLDYEBEDLLYWDVDSLQRIERST 2579  
QY 314 ---VD---W---Q-----S-----N----- 319  
Db 2580 LTGVDRREVIVNAAHAFGLTLYGQYIYWDLTQRIYRANKYDGSQIAMIATNLLSQPRG 2639  
QY 320 -NT-----F----- 322  
Db 2640 INTVKNQKQCNPCPEQNGGSHICAPGNGAECQCQHEGNYLANNRKHCIUVNGER 2699  
QY 323 --AS---CS---TD-M---C-IH-----V----- 333  
Db 2700 CGASSFTCSNGRCISBEWKCDNDNDGDSDEMESVCALHTCSPTAFTCANGRCVQYSYR 2759  
QY 334 CKL-----G---QD-----R---P---I----- 342  
Db 2760 CDYINDGDSDEAGCLFRDCNATTFMCNRRRCIPREFICNGVDNCHDNNTSDEKXCPD 2819  
QY 343 ---K-----T-----FQ----- 346  
Db 2820 RYCQSGYTKHNSNICIPRYVLCGDNDCGNSDENPTYCTHTCSSEFQASGRICIPQ 2879  
QY 347 ---GH---T----- 349  
Db 2880 HWYCDQETDCFDASDEPASCGRHSERTCLADEPKDGGRCIPSEWICDGDNDGDMDEDK 2939  
QY 350 ---N---E---VNA---IK---W----- 357  
Db 2940 RHQONQNGSDFLCVNDRPPDRRCIPQOSWVDCGDVDCDGDYDENONCTRRTCSENEFT 2999  
QY 358 ----- 357  
Db 3000 CGYGLCIPKIFRCDRINDCGYSDERGCLYQTCQOQFTCQNGRCISKTFVCDENDCGD 3059  
QY 358 --D---PT-----G---N---L----- 363  
Db 3060 GSDELMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCNHLDDCLDNDSEKGGINECHDPSI 3119  
QY 364 ---LAS---CS-----D-D---M----- 371  
Db 3120 SGCDHNCTDLTSFYCSRPGYKLMSDKRTCDVIDECTEMPFVCSOKCENIVGSIYCKCA 3179  
QY 372 ---T---L-----L-----K 374  
Db 3180 PGYLREPDGKTCEQNSNIEPYLIFSNRYLRNLITIDGYFVSLLEGLDNVVALDFRVEK 3239  
QY 375 -----I-----W-----S--- 377  
Db 3240 RLYWIDTQQVIERMFLNKTNETIINHRLPAESLAVDWVSRKLYWLDARLDGLFVSDL 3299  
QY 378 ---M-KQ---D---N---C-----VHDLOO-----H-----NK- 392  
Db 3300 NGGHRMLAQHCVDANNTFCFDPNPRGLALH--PQYGLVWADWGHRAIYGRVGMGTNKS 3357  
QY 393 ---E---I-YT-----IK----- 398  
Db 3358 VIISTKLEWPNGTIDYTNLLYWAHAHLGYIEYSDLEGGHHRHVVYDGLPHFPFAITIFE 3417  
QY 399 ---W-----S-----P-T---G--- 403  
Db 3418 DTIYWDNTRTVEKNGKYDGSNRQTLVNTTHRPFDIHVVHPYRQPIVSNPCGTNNGCS 3477  
QY 404 ---PG-----T---NN----- 408  
Db 3478 HLCLIXPGGKFTCECPDDPFTLQLSGSTVCMPCWCSSTQFLCANNEKCIPIWKCQDQKD 3537  
QY 409 ---P---NA---N-----LML-----AS 417  
Db 3538 CSDGSDALALCPORFCELQFQCGDNGCTSPQTLCAHQNCPDGSDDEDL-L-CENHHCHDS 3596  
QY 418 ---A---S---F-----DS---T---VR---L---W---DV 429



Db 3597 NEWQCANRCIPESWQCDTFNDCEDNDESSHCASRTCRPGQFRANGRCIPQAWKCDV 3656  
QY 430 --DRG-----I--CI--H-----T--L-----438  
Db 3657 DNDGHDSEPIECSSAHLCDNFTEFSKTYRCIPKAVCNVDDCDNDSDGCEE 3716  
QY 439 -T-----K--H-----Q-----E--P-----V-----445  
Db 3717 RTCHPVGDFRCKNHCIPLRWQCDGNDGDNDEENCAPECTESEFRVCNQICPSRW 3776  
QY 446 -----Y-----S--V-----A-----F 450  
Db 3777 ICDHYNDCGNSDRDCMRTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCPTRF 3836  
QY 451 SPDGRYL-AS-----GS-----F-----462  
Db 3837 -PDGATQATMFCKNHVICPPYWKCDGDDGDBELHLCLDVPKNSPFRCDNNR 3895  
QY 463 -----D-----K-----CV-H-I-----468  
Db 3896 CIYSHEVCNVDGDTDEBHCRTKPTKPTCEYKCGNGHCIPHDNVCDADDDCGD 3955  
QY 469 W-----N-----Q--T-----G-----A-----L-V-----477  
Db 3956 WSDLGCKGKERTCAENICEONCTQNEGFFCSCCTAGFTNVFDRTSCLDINECEQFG 4015  
QY 478 -----H-----SV-----R-----481  
Db 4016 TCPQHCNTKGSYECVACDFTSMRDPGRKCAEGSSPLLLLPDNNVRKYNLSRERS 4075  
QY 482 -----G-----T-----G-----G-----I--FE-----V-C-- 490  
Db 4076 EYLQDEEYIOAVDYDMDPEDIGLSVYVYTVRGESRFGAKRAVINFESGRNLLVQEV 4135  
QY 491 -----W-----N-----AA-----494  
Db 4136 LKLKVMQPDGIAVWGRHYNSDVKNKRIEVAKLDGRYKWLISTDLDPQAAIAVNP 4195  
QY 495 -----G-----G-----D-K-V-- 498  
Db 4196 LGLMFTWCKBPKXSAMNGEDRNLVLPEDLGNPTGLSIDYLNNDRIYWSDFKEDVIE 4255  
QY 499 -----G-----A-SA-----S-----D-----G-----505  
Db 4256 TIKYDGTDRVIAKEANPVSLLIFEDQLYWISKEGEVWKQNGKKEKTLVNPWL 4315  
QY 506 -----SV-----C-V-----L-----D-----511  
Db 4316 TVRIFHQLRYNKSVPNLCKQICSHLCLLRPGGYSCACPOGSSFIEGSTTECDAAIELPI 4375  
QY 512 -L-----R-----K 514  
Db 4376 NLPPPCRMHGGNCYFDETLDPK 4398

RESULT 7

US-08-652-877-90  
; Sequence 90, Application US/08652877  
; Patent No. 6187548  
; GENERAL INFORMATION:  
; APPLICANT: Akerstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Rask, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjaln, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
; TITLE OF INVENTION: Thereof and DNA Encoding Same  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd., 3C43

; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.5.1  
; SOFTWARE: Word 6.0 (Patentin)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/652,877  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/15203  
; FILING DATE: 22-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/344,836  
; FILING DATE: 23-NOV-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/487,314  
; FILING DATE: 07-JUNE-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitzky, Martin  
; REGISTRATION NUMBER: 29,699  
; REFERENCE/DOCKET NUMBER: A1355E-US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-454-3816  
; TELEFAX: 610-454-3808  
; INFORMATION FOR SEQ ID NO: 90:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4655 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-652-877-90

Query Match 70.1%; Score 2566.6; DB 3; Length 4655;  
Best Local Similarity 9.8%; Pred. No. 5.1e-32;  
Matches 433; Conservative 57; Mismatches 19; Indels 3894; Gaps 381;

QY 1 M-----S-----I-----S--SDE- 7  
Db 1 MDRGPAVACTLLALVALCLAPASGQECDSAHFRGSGHCIPADWRCDGTGKDCSDDADEI 60  
QY 8 ---VN---P---L---V-----12  
Db 61 GCAVVTCCQGYFKQSEGCIPSSWVCDQDQDCDSDERQDCSQTCSHQITCSNGQC 120  
QY 13 ---YR---L---Q-----E--S 19  
Db 121 IPSEYRCDHVRDCPDGADENDCOYPTCEQLTCNGACYNYSQKDWKDCRSDSDEINCT 180  
QY 20 ---G-FS-----H-----S-A-F-T---F---G-I--- 30  
Db 181 EICLHNEFSCGNGECIPRAVVCDDHNDQDGSDEHACNYPTCGGYQFTCPSGRCIYQNWV 240  
QY 31 -----K-----S-----H-----IS-----Q--SNI 39  
Db 241 CDGEDCKDNGDBDGCESGPHDVHKCSPREWSCPSGRCISIIYKVCDDGILDGCPGREDEN- 299  
QY 40 N--G-----AL-----VP-----PAALI-----S-----I--I-- 53  
Db 300 NTSTGKYSMTLCSALNCYQCHETPYGACFCFPGYIINHNSRVCVEFDQCIQWICD 359  
QY 54 QK-----G--L---QY-----VE 61  
Db 360 QKCESRPRHLCHCEBGYILERQGYCKANDSPGEASIIFNSGRDLLIGDIHGRSFRILVE 419  
QY 62 -----A-----E--VSI-----N 67  
Db 420 SQNRGVAVGVAFHYHLQRFVWTDVQNKVFSVDINGLNIEQLVNSVETPENLAVDWNV 479

Qy 68 ---E---DG---TL-----72  
Db 480 KIYLVETKVRIDMWNLDGSRVTLITENLGHPRGIAVPTVGYLFESDNESLSGEPKLE 539  
Qy 73 -F-DG-R-----P-----IE-----S 80  
Db 540 RAFMDGSRNRKDLVKTKLGPAGVTLDMISKRVYVWDSRFDYIETVTVYDGIQRKTVVHGS 599  
Qy 81 L-----SL-----I-----D-----AVM-----88  
Db 600 LIPHPGVSLEFGQVFTDWTMKAVALKANKFTETNPQVYQASLRPYGVTVYHSLROPYA 659  
Qy 89 -P-D-----V-V-----Q-T-R-----Q-----Q-AY 99  
Db 660 TNPCKDNNGGCEQVCLSHRTDNDGLFRCKTFGFOLDTDERHCIAVQNFIFSSQVAI 719  
Qy 100 R-----D-K-----L-AQ-----Q-----Q-A 108  
Db 720 RGIPFTLSTQEDVMVPVSGNPSFPVGDIDFAQDSTIFFSDMSKHMIFKQKIDGTGREILA 779  
Qy 109 A-----A-----A-----A-----A-----A--112  
Db 780 ANRVENVESLAFDWISKNLWYTDSHYKSI SVMRLADKTRTVVQYLNPNRSVVVHPAGY 839  
Qy 113 -----A-----A-----A-----AA-----S--119  
Db 840 LFFTDWFRPAKIMPAWSGSHLLPVINTILGWPNGLAIDWAASRLYWVDAYFDKIEHSTF 899  
Qy 120 -----Q-----Q-----Q-----Q-----Q-----120  
Db 900 DGLDRRLGHTEQHTPFLAIFGEHLFFTDWRLGAIRVRKADGGEMTVIRSGIAVILH 959  
Qy 121 -----Q-GS-A-----KNG-----EN-----TA-N-----G-----133  
Db 960 LKSYDVNIQTGSNACNQPHTPGDCSHCFPFVFNFORVCGPYGMRLASNLHLCFEGDPTN 1019  
Qy 134 E-E-----N-----GA-H-----T-----T-----140  
Db 1020 EPPEQCGLFSPCKNGRCVNYLVDGDDCHNSDEQLCGTLNNTCSSAFTCGHGEC 1079  
Qy 141 I-A-----N-----H-----T-DMME-----V-D-----GD-----154  
Db 1080 IPAHRWCDCRNDKVDGSDHNCPTAPASCLDTQYTCNDHQCSKNWCVCDTNDGCGSD 1139  
Qy 155 -----VEI-P-----PN-----K-A-----162  
Db 1140 EKNCNSTETCQPSQFNCNHRCDLSFVCDGDKDCVDGSDVGVCLNCTASQPKASGDK 1199  
Qy 163 ---V-----VL-----R-G-H-----E-----169  
Db 1200 CIGVTRNRCGVDCSDNSDEAGCPTRPFGMCHSDFEQCEGICIPNFWECDGHPDCLYG 1259  
Qy 170 S-E-----V-----F-----IC-AW-----N-----P-----179  
Db 1260 SDEHNACVPKTCPSYFHCNCGNCIHRWLCDRNDGDMSEKDCPTQPFRCPSWQWC 1319  
Qy 180 -----V-----V-----SD-----182  
Db 1320 LGHNICVNLVSVCDGIFDCPNGTDESPLCNGNSCDFNGGCTHECVQEPFGAKCLCPLGF 1379  
Qy 183 LLA-----S-----S-----G-----189  
Db 1380 LLANDSKTCEDIDECILGSCSQHCYNMRGSRFCSDTGYMLSSDGTCKVTASESILLL 1439  
Qy 190 -----DS-T-----A-----RI-----W-----N-----197  
Db 1440 VASQNKIADSVTSQVHNIYSLVENGSIYVAVDFDSISGRIFWSDATQGTWSAFQNGTD 1499  
Qy 198 -----LSE-----N-S-----T-----S-----GS-----T-Q-LVL 211  
Db 1500 RRVVDSIIILTETIADWGRNLYWTDYALETIEVSKIDGSHRTVLISKNLTNPRGLAL 1559  
Qy 212 --R-----H-----213

Db 1560 DPMNEHLLFWSDGWHGPRIERASMDGSMRTVIVQDKIWPFCGLTIDYPNRLLYFMDSYL 1619  
Qy 214 ---C-----IR-----E-----GG-----219  
Db 1620 DYMFCDYNGHRRROVIASDLIIRHPYALTLPEDSVYWTDRATRRVMRANKWGHGQSVV 1679  
Qy 220 -----O-----O-----220  
Db 1680 MYNIQWPLGIVAVHPKQNSVNPCAFSRCSHLCLSSQGPHPYSCVPCSGWLSLSPDLLN 1739  
Qy 221 ---D-----V-----P-SN-K-----DV-----228  
Db 1740 CLRDDQFLITVRQHIIIFGILNPEVKSNDAMVPIAGIQLNGLDVEFDABEQIYWVENPG 1799  
Qy 229 ---T-----S-LDW-----N-S-E-----G-----T 238  
Db 1800 EIHVRKTDGTRNRTVFASISMGVGSMLALDWISRLNLYSTNPTQSIETLHGDIRYRKT 1859  
Qy 239 LLA---T-G-----S-Y-----Y-----DG-----F-----248  
Db 1860 LIANDGTALGVPPIGITVDPARGLYSDQDGTSDGVPKIASANMDGTSVKTLFTGNLE 1919  
Qy 249 -----A-----R-I-----W-----T-----253  
Db 1920 HLECVTLDIEEQKLYWAVTGRGVIERNVDGTRMILVHQLSHPMGIAVHDSFLYVTDEQ 1979  
Qy 254 -K-D-G-N-L-ASTL-G-Q-H-----K-----266  
Db 1980 YEVIERVADKATGANKIVLRDNPVNLRLGLQVYHRRNAESNGSCNNMNACQICLPVPGG 2039  
Qy 267 ---G-----P-----I-FALK-----273  
Db 2040 LFSCACATGKLPNDRSCSPNSFVIVMSLGAIRGFSLELSDHSETMPVACQGRNALH 2099  
Qy 274 ---W-----N-----K-----K-----GN-279  
Db 2100 VDVVSSGFIYCDPSSVASDIAIRIKPDGSSLMNIVTHGIGENGVRGIAVDWAGNL 2159  
Qy 280 -----F-----I-----281  
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTVMRPHRVDPKRYLFWADYQORPKIERSF 2219  
Qy 282 ---L-S-----A-G-V-D-----287  
Db 2220 LDCNTRVLVSEGIPTPRGLAVDRSDGYVYVWDDSLDIARIPINGENSEVIRYGRYPT 2279  
Qy 288 ---K-T-T-I-I-W-----D-----294  
Db 2280 PYGITVPENSIWVDRNLKKIIOASKEPEPTEPTTVIRDINWLDRDVTIFDKQVQPRSPA 2339  
Qy 295 ---A-H-T-----G-----E-A-----300  
Db 2340 EVNNPNLENNGGCSHLCPALPGLHTPKDCAFGLQSDGKNCAISTENFLIPALSNSLR 2399  
Qy 301 -K-Q-----Q-F-----P-305  
Db 2400 SLHLDPENHSPPTQINVERTVMSLDYSDVSDRIYFTQNLASGVQISVATLSSGHTPT 2459  
Qy 306 -----F-----306  
Db 2460 VIASGIGTAGIAFDWITRRIYVDYNQWINSMAEDSGSNRTVIARVPKPRAIVLDPQCG 2519  
Qy 307 ---H-----S-AP-ALD-----313  
Db 2520 YLYAWDWDTHAKIERATLGNFRVPIVNSLWPSGLTLDYEDLLYVWDASLQRIERST 2579  
Qy 314 ---VD-----W-Q-----S-N-----319  
Db 2580 LGVDEVIVNAVHAFGLTYLQYIYWTDLTYQRIYRANKYDGSQIAIANTLLSQPRG 2639  
Qy 320 -NT-----F-----322

Db 2640 INTVKNQKOCNNPCQFNGGSHICAPGNGAEOCPHEGNWYLANNRKHCIVDNGER 2699  
Qy 323 --AS--CS-----TD-M--C-IH-----V----- 333  
Db 2700 CGASSFTCSNGRCISBWKCDNDNDGDSDESVCAHTCSPTAFTCANGRCVQYSYR 2759  
Qy 334 CKL-----G-QD-----R-P-I----- 342  
Db 2760 CDYVNDGDSDEAGLCFRDCNATTEPMCNRRCPREFICNGVDNCHDNTSDEKNCPD 2819  
Qy 343 -----K-----T-----PQ----- 346  
Db 2820 RTCOSYTKCHNSNICIPRYLDCGNDGNSDENPTYCTTHTCSSEFQCSAGRCIPQ 2879  
Qy 347 -----GH--T----- 349  
Db 2880 HWYCDQETCFDASDEPASGHSERTCLADEFKDGGRCIPSEWICGDNDCGMSDEK 2939  
Qy 350 -----N-----E--VNA-----IK--W----- 357  
Db 2940 RHQONQNCSDSEFLCVNDRPPDRRCIPQSMVCDGDVDCDGYDENQNCRTTRTCSNEFT 2999  
Qy 358 -----PT-----G-----N-----L----- 357  
Db 3000 CGYGLCIPKIFCRDHNDGYSDBRGCLYQTCQOQFTCQNGRCISKTFVCDENDDCG 3059  
Qy 358 -----D-----PT-----G-----N-----L----- 363  
Db 3060 GSDLMHLCHTPEFTCPPEFKCDNGRCIEMKLCNHLDCDLSDEKGGCINECHDPSI 3119  
Qy 364 -----LAS--CS-----D-D-----M----- 371  
Db 3120 SGCDHNCTDLTSFYCSRGYKLSDKRCTVDIDECTEMPFVCSKCNVIGSYICKCA 3179  
Qy 372 -----T-----L----- 374  
Db 3180 PGYLRPDGKTCRQNSIEPYLFSNRYLRNLTDIGYFYSLEGLDNVVALDFRVEK 3239  
Qy 375 -----I-----S-----W-----S----- 377  
Db 3240 RLYWIDTQVIERMFLNKTKETIINHRLPAESLAVDWVSRKLTWLDARLDGLFVSDL 3299  
Qy 378 -----M-KQ--D--N--C-----VHDLOQ-----H-----NK-- 392  
Db 3300 NGCHRRMLAQHCVDANNTFCFNDPRGLAH--PQYGLYWADWGHAYIGRVGMDGTNKS 3357  
Qy 393 -----E-----I-YT-----IK----- 398  
Db 3358 VIISTKLEWPNGITIDYTNLLYWADAHGLYIEYSDLEGHHRHTVYDGLPHFPFAITFE 3417  
Qy 399 -----W-----S-----P--T--G-- 403  
Db 3418 DTIYTDWTRTVEKNGKYGNSRQTLVNTTHRPDIHVHPYRQPIVSNPCGTTNGGCS 3477  
Qy 404 -----PG-----T-----NN----- 408  
Db 3478 HLCILKPGGKFTCECPDDRTILQSGSTYCMPCSSQPLCANNEKCIPIWKKDQXD 3537  
Qy 409 -----P--NA--N-----IML-----AS 417  
Db 3538 CSDGDELALCPQFRLGQFCSDGNCSTSPOTLCNAHQCPDGSDEDL-LCENHCHDS 3596  
Qy 418 -----A-----S-----P-----DS-----T-----VR-----L--N--DV 429  
Db 3597 NEWQCANRCIPESWQCDTNDCEDESDSDSSHCASRTCPGQFRANGRCIPQAWKCDV 3656  
Qy 430 -----I--CI--H-----T--L----- 438  
Db 3657 DNDGDSHDEPIECMSSAHLCDNFTFECKTYRCIPKWAVCNVDDCDNDSDEQCEE 3716  
Qy 439 -T-----K--H-----Q-----E--P-----V----- 445  
Db 3717 RTCHPVGDFCKNHHCIPLRWQCDGQNDGNSDEENCAPECTSEFRVCVNOQCIPSRW 3776

Qy 446 -----Y--S--V-----A-----F 450  
Db 3777 ICDHYNDGNSDERDCMRKTCHEPYFQCTSGHCVHSELKCDGSADCLDASDADCPTRF 3836  
Qy 451 SPDGRYL-AS-----GS-----F----- 462  
Db 3837 -PDGAYCOATMECKKNHCIPPYWKCDGDDDCGSDGDEELHCLDVPKSPNFRCDNNR 3895  
Qy 463 -----D-----K-----CV-H-I----- 468  
Db 3896 CIYSHEVGVNDDCGDGTDETEHCKRPTKPCTEYKCGNGHCIPHDNVCDADDCGD 3955  
Qy 469 W-----N-----T-----Q--T-----G-----A-----L--V----- 477  
Db 3956 WSDELCGNKGRKTCACENICEQNCTQLNEGGFICSTAGTETVFDRTSLDINECEQFG 4015  
Qy 478 -----H-----SY-----R----- 481  
Db 4016 TCPQHCRNTKGSYECVACDGFTSMRPGKRCAAEGSSPLLLLPDNNVRIRKYNLSERFS 4075  
Qy 482 -----G-----T-----G-----G-----I--EE-----V--C-- 490  
Db 4076 EYLQDEBYTQAVDYDWDPEDIGLSVVYTVRGEGRFGAKRAYIPNFESGRNMLVOEVD 4135  
Qy 491 -----W-----N-----AA----- 494  
Db 4136 LKLYVMQPDGIAVDWVGRHIYWSDVKNKRIEVAKLDGRYKRWLISTDLDQPAIAVNP 4195  
Qy 495 -----G-----D-----K--V-- 498  
Db 4196 LGLMFTWKGKPLESAWNGEDRNILVFDLGMPTGLSIDYLNNDRIYWSDFKEDVIE 4255  
Qy 499 -----G-----A-SA--S--D-----G----- 505  
Db 4256 TIKYDGTDRRTAKEANNPYSLDIFEDQLYWISKEKEVWKNQKFGGKKEKTLVNPWL 4315  
Qy 506 -----SV--C--V-----L-----D----- 511  
Db 4316 TVRIFPHQLRYNKSVPNLCKQICSHLCRLRPGYSCACPGQSSFIEGSTTECDAAIELPI 4375  
Qy 512 -L-----R-----K 514  
Db 4376 NLPPPCRCMHGNCYFDETDLPK 4398

## RESULT 8

US-08-652-877-86  
; Sequence 86, Application US/08652877  
; Patent No. 6187548  
; GENERAL INFORMATION:  
; APPLICANT: Akerstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Raek, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjalms, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
; TITLE OF INVENTION: Thereof and DNA Encoding Same  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd., 3C43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.5.1  
; SOFTWARE: Word 6.0 (Patentin)

Qy	1 M----	-S-----I-----S--SDE- 7
Dd	1 MDRGPAAVACTLLALVACLAPASGOBCDSAHFRCSGHGHCIPADWRCDGTKDCSDDAEI 60	:
Qy	8 ---VN---F---L---V-----12	:
Dd	61 GCAVTCQQGYFKCQSGQCIPSSWVCDDQDGDGDERQDCSQSTCSSHQITCSNGQC 120	:
Qy	13 ---YR-----Y---L-----Q-----E--S 19	:
Dd	121 IPSEYRCDHVRDCPDGADENDCQPTCEQLTCDNGACVNTSOKCDWKVDCRDSSDEINCT 180	:
Qy	20 ---G-FS-----H-----S-A-F-T-----P-G-I-----30	:
Dd	181 EICLNHFSCNGECIPRAYVCDHDNDQDGSDHACNYPTCGGYQTCPSGRCIYNQWV 240	:
Qy	31 -----K---S---H-----IS-----Q--SNI 39	:
Dd	241 CDGEDCKNGDEDGESGPHVKCSPREWSCPSGRCSIYKVCDGILDCPBREDN- 299	:
Qy	40 N---G-----AL-----VP-----PAALI---S-----I-I-- 53	:
Dd	300 NTSTGKYCSMTLCSALNCQYCHETPYGGACFCPPGYIINHDSRTCVFPDDCIWGICD 359	:
Qy	54 QK-----G-L---QY-----VE 61	:
Dd	360 QKESRPRHLCHCEGYILLERGOYCKANDSFGEASIFNSGRDLLIGDIHGSRFILVE 419	:
Qy	62 ----A-----E-----VSI-----N 67	:
Dd	420 SQNRGVAVGAFYHLORFWTVDVQNKFVSVDINGLNIQVLNVSVETPENLAVDWNVN 479	:
Qy	68 ----E-----DG---TL-----72	:
Dd	480 KIVLVETKNRIDMWNLDGSRYVTLITENLGHPRGIAVDPVTGYLFFSDWSLSGPKE 539	:
Qy	73 --F-DG-R-----P-----IE-----S 80	:
Dd	540 RAFMDGSKDLVTKLGWPAGVTLDMISKRVYWDSRFDYIETVTDVGIOKTVVHGGS 599	:
Qy	81 L-----SL---I---D---AVM-----88	:

Db 1680 MYNIQWPLGIVAVHPKQPNVNPFCARSRCSHLCLLSSQGHFHYSCVCPGWSLSPLDLN 1739  
QY 221 ---D---V---P---SN---K---DV--- 228  
Db 1740 CLRDDQFLITVRQHIIIFIGISLPEVKSNDAMPIAGIQLGDLVEFPDPAEQIYIWNENPG 1799  
QY 229 ---T---S---LDW---N---S---E---G---T 238  
Db 1800 EIHVRKTDGTRNVTFASISMVGFPMNLALDWISRLNLYSTNPTQTSIEVLTGLHGDIRVKT 1859  
QY 239 LLA---T---G---S---Y---DG---F--- 248  
Db 1860 LIANDGTALGVGPIGIVDPARKLYWSQDGSVPAKIASANMDGTSVKTLFTGNLE 1919  
QY 249 ---A---R---I---W---T--- 253  
Db 1920 HLBCTVLIDIEQKLYNAVTGRGVIERNVGDTRMILVHQLSHPWGLIAVHDSFLYYTDEQ 1979  
QY 254 ---K---D---G---N---L---ASTL---G---Q---H---K--- 266  
Db 1980 YEIVERVKATGANKIIVLRDNVFNLRGLQVYHRRNAEASSNGCSNNMNAQQICLPVPGG 2039  
QY 267 ---G---P---I---PALK--- 273  
Db 2040 LFSCACATGFKLAPDRNRSCTPYNFIVVMSLSAIRGFSLSHSETMVPVAGQGRNALH 2099  
QY 274 ---W---N---K---K--- 279  
Db 2100 VDVVSSGFIYWCDFSSVASDNAIRIKPDGSSLMNIVTHGIGENGVRGIAVDWAGNL 2159  
QY 280 ---I---F---I--- 281  
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTVDMPRHIVVDPKNRYLFWADYGQPKIERSF 2219  
QY 282 ---L---S---A---G---VD--- 287  
Db 2220 LDCTNRTVLVSEGIIVTPRGIAVDRSDGYVYVVDSDLDIARIINGENSEVIRYGRYPT 2279  
QY 288 ---K---T---T---I---I---W---D--- 294  
Db 2280 PYGITVFENSIIVDRNLKIKIQAQKEPENTEPPTVIRDNINWLRDVTIPDKQVQPRSPA 2339  
QY 295 ---A---HT---G---E---A--- 300  
Db 2340 EVNNNPCLENGGCSHLCPALPGLHPTKCDCAFGLQSDGKNCAISTENFLIALNSLR 2399  
QY 301 ---K---Q---Q---F---P--- 305  
Db 2400 SLHLDPENHSPPQTINVERTVMSLDYSDRIYFTQNLASGVGQISYATLSSGIHTPT 2459  
QY 306 ---F--- 306  
Db 2460 VIASGITADGIAFDWITRIIYSDYLNQMINMAEDGNSRTVIARVPKRAIVLDPQCG 2519  
QY 307 ---H---S---AP---ALD--- 313  
Db 2520 YLWADWDTHAKIERATLGNFRVPIVNSSLVMPGSLTLDYEEDLLYWDASLQRIERST 2579  
QY 314 ---VD---W---Q---S---N--- 319  
Db 2580 LTGVDRREVINAHAFLGLTYGYIYWTDLTYRIYRANKYDGSQGIAMTTNLLSQPRG 2639  
QY 320 ---NT---F--- 322  
Db 2640 INTVVRNQKQCCNPNCEQFNGGCSHICAPGNGAECQCPEHGNWYLANNRKHCIVDNGER 2699  
QY 323 ---AS---CS---TD---M---C---IH---V--- 333  
Db 2700 CGASSFTCSNGRCISBEWKCDNDGCGSDENESVICALHTCSPTAFTCANGRCVQYSYR 2759  
QY 334 CKL---G---QD---R---P---I--- 342  
Db 2760 CDYINDCGGSDBAGCLFRDCNATTEFMCNRRRCIPREFICNGVDNCHDNTSDEKNCPD 2819

QY 343 ---K---T---FQ--- 346  
Db 2820 RTCQSGYTKCHNSNICIPRVYVLCGDNDCGDSDENPTYCTTHTCSSEFQCTSGRCIPIQ 2879  
QY 347 ---GH---T--- 349  
Db 2880 HWYCDQETDCFDASDEPASCGRHSERTCLADEFKCDGGRICIPSEWICDNDNCGDMSDEDK 2939  
QY 350 ---N---E---VNA---IK---W--- 357  
Db 2940 RHOCQONQSDSFLCVNDRPPDRRCIPOSWCVGCDVDCDGYDENQNCNTRTTCSENEFT 2999  
QY 358 --- 357  
Db 3000 CGYGLCIPKIFRCDRHNDGDISDERGCLYOTCQOQFTCQNGRCISKTFVCDENDCGD 3059  
QY 358 ---D---PT---G---N---L--- 363  
Db 3060 GSDELMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCNHLDDCLDSDKGGGINECHDPSI 3119  
QY 364 ---LAS---CS---D---M--- 371  
Db 3120 SGCDHNTDTLTSFYCSCRPGYKMGDKRTCDIDECTEMPFVCSQKCNVIGSYICKCA 3179  
QY 372 ---T---L--- 374  
Db 3180 PGYLREPKDTCRQNSNIEPYLIFSRYLYRLNLTIDGYFYSLLLEGLDNVVALDFRVEK 3239  
QY 375 ---I---W---S--- 377  
Db 3240 RLWIDTQROVIERMFLNKTNETIINHRLPAAESLAVDMVSRKLYWLDARLDGLFVSDL 3299  
QY 378 ---M---KQ---D---N---C---VHDLOQ---H---NK--- 392  
Db 3300 NGHREMLAQHCVDANNTFCFONPRGLAH---PQGYLYWADWGHRAYIGRVGMDGTNKS 3357  
QY 393 ---E---I---YT---IK--- 398  
Db 3358 VIISTKLEWPGITIDYTNLDLWADAHLYIEYSDLEGHHRHTVVDGALPHFAITIFE 3417  
QY 399 ---W---S---P---T---G--- 403  
Db 3418 DTIYTDWTRTVKEKNGKYGNSRQTLVNTHTRPFDIHVYHPYRQPIVSNPCGTNNGCS 3477  
QY 404 ---PG---T---NN--- 408  
Db 3478 HLCLIKPGKGFTCECPDDFRTLQLSGTYCMPMCSSTQFLCANNEKCIPIWKKCDGQKD 3537  
QY 409 ---P---NA---N---LML---AS 417  
Db 3538 CSDGSEALALCPQRCRLGQFCQSDGNCSTSPQTLCAHQNCPDGSDEDL---LCENHHCD 3596  
QY 418 ---A---S---F---DS---T---VR---L---W---DV 429  
Db 3597 NEWQCANKKECIPESWQCDTFNDCEDNDESDSHCASRTCPGQFCRANGRCIPIQAMKCDV 3656  
QY 430 ---DRG---I---CI---H---T---L--- 438  
Db 3657 DNDGSDHSDPEIEECSSAHLCDNFTEFSCKTNYRCIPKWAQVGVDDCDRDNSEOGCEE 3716  
QY 439 ---T---K---H---Q---E---P---V--- 445  
Db 3717 RICHVPVGDPRCKNHHCIPLRWQCDGNDGNDSDENECAPRECTESEFRVCNQCIPSRW 3776  
QY 446 ---Y---S---V---A---F 450  
Db 3777 ICDHYNDGNDSDRDERCEMRTCHPEYFQCTSGHCVHSELKCDGSADCLDASDEADCPTRF 3836  
QY 451 SPDGRYL---AS---GS---F--- 462  
Db 3837 -PDGAYCATMFECKNHVICIPPYWKCDGDDGCGSDEBELHCLDVPNCSPNFRCDNNR 3895

QY 463 -----D-----K-----CV-H-I----- 468  
Db 3896 CIYSHEVNGVDCGDTDETEBHRKPTPKCTEYKCGNGHCIPHDNVCDADDGCD 3955  
QY 469 W-----N-----Q-T-----G-----A-----L-V----- 477  
Db 3956 WDELGNKKEKTCENICEQNTQINBEGFICSTAGPENVFDRTSCLDINECEQFG 4015  
QY 478 -----H-----SY-----R----- 481  
Db 4016 TCPQHCNRTKGSYECVCADGFTSMRPGKRCACAGSSPILLLPDNVRIKYNLSERFS 4075  
QY 482 -----G-----T-----G-----G-----I-EE-----V-C----- 490  
Db 4076 EYLQBEYIOAVDYDWDPDIGLSVVYTVRGESRFGAIKRAYIINFESGRNLVQEV 4135  
QY 491 -----W-----N-----AA----- 494  
Db 4136 LKLKYMOPDGIADVGRHIYWSDVKNKRIEVAKLDGRYRKWLSTDLDPQAAIAVNP 4195  
QY 495 -----G-----G-----D-K-V----- 498  
Db 4196 LGLMFTDWCKEPKXESAMNGEDRNILVPEDLGWPTGLSIDYLNNDRIYWSDFKEDV 4255  
QY 499 -----G-----A-SA-----S-----D-----G----- 505  
Db 4256 TIKYDGTDRVIAKAMNPSLDIFEDQLYWISKEGEVWKQKPGQKKEKTLVNPWL 4315  
QY 506 -----SV-----C-V-----L-----D----- 511  
Db 4316 TVRIFHQLRYKNSVNPNCQICSHLCLLRPGGYSCACPOGSSFIEGSTTECDAAI 4375  
QY 512 L-----R-----K 514  
Db 4376 NLPPPCRMHGNCYFDETLDPK 4398

## RESULT 9

US-08-718-388-9  
; Sequence 9, Application US/08718388  
; Patent No. 6271362  
; GENERAL INFORMATION:  
; APPLICANT: MORIKAWA, MINORU  
; APPLICANT: HARADA, NAOKI  
; TITLE OF INVENTION: GENE ENCODING IgG Fc REGION-BINDING  
; TITLE OF INVENTION: PROTEIN  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/718,388  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 0230-111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5405 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-718-388-9

Query Match 69.8%; Score 2553.6; DB 3; Length 5405;  
Best Local Similarity 9.6%; Pred. No. 2.1e-31;  
Matches 412; Conservative 80; Mismatches 17; Indels 3764; Gaps 356;

QY 1 M-----S-I-----SS-D-EVN-----ELV-----YR-----Y----- 15  
Db 1 MGLMSWILWAGATLLMGLTQEASVDLK-NTGRBEELTAFLONYQLAYS KAYPRLLISS 59  
QY 16 -----L-Q-----ES-----G-F-S-H----- 23  
Db 60 LSESPASVILLSQADNTSKVTVPGESVWVNISAKAEMIGSKI FQHAVVHSDVAISVQ 119  
QY 24 -----SA-----F-----T 27  
Db 120 ALNAKPDTAELTLRPIQALGTEYFVLTTPPGTSARNVKEFAVVAGAAGASVSTLKGSVT 179  
QY 28 F-----G-----I-----K-----S-----HI- 34  
Db 180 FNGKFPAGDVLRLTLPQVNVQLOSSVDLSGSKVTASSPVAVLSGHSCAQKHHTCNHV 239  
QY 35 -----SQS-----N-----I-----S----- 39  
Db 240 EQLLPTSANGTHVVPVTLASQSRYDLAFVVASQATKLTYNHGGITGSRGLQAGDVVEFEV 299  
QY 40 -----N-----GAL-V-P-----P-AAL-----I-S----- 51  
Db 300 RPSWPLYLSANVGIOVLLFETGAIRNEVTDPYLVLPDVAAYCPAYVVKVSPGCEGVAL 359  
QY 52 II-Q-K-----G-LQY-V-E-REV-S-I-----N----- 67  
Db 360 VVAQTKAISGLTIDGHAVGAKLTWEAVPGSEFSYAEVELGTADMIHTAETNLGLLTG 419  
QY 68 --E-----D-G-T-L----- 72  
Db 420 LAKAIGYATAADCGRTVLSVPBSCBGMQCAAGORCQVVGKAGCVAESTAVCRAQGDPH 479  
QY 73 ---FDGR-----P---IE----- 79  
Db 480 YTTFDGRRYDMMGTCSYTMVELCSEDDTLTLPFAFSEAKNEHRSRRRVYVGLVTVRAYSHS 539  
QY 80 -SLS-----LID-----AV-----M 88  
Db 540 VSLTRGEVGFVLVDNQRSLPVSLSRGLRVYQSGPRAVVELVGLVTVYDWDQLALS 599  
QY 89 P-----D-V-----V-----Q----- 93  
Db 600 PARFQDVQVCGLCNGYNGDPADDFLTDPGALAPDAVEFASWSKLDGDYLCEDCGCQNNCPA 659  
QY 94 -TR-Q-QAY---R-----D-KLA-----Q-----Q----- 106  
Db 660 CTFGQAQHYEGDRLCGMLTKLDGFFAVVCHDTLDPRLFEQVYDLCVVGGERLSLCRGLS 719  
QY 107 ---QA-----A-----A-----A-----AAA----- 114  
Db 720 AYAAQACLELGISVDWRSPANCPLSCFANSRYELCGPACPTSCNGAAAPSNCSGRPCVEG 779  
QY 115 -----A---A---AAS-----Q-Q---G-S-A-K-----NGEN---T----- 130  
Db 780 CVCLPGFVAGSGGACVPASSCGCTFQGLQAPGOEVMWDELCCORRCTCNGATHQVTCRDKQ 839  
QY 131 ---ANGE-----EN-----GA---H-----T---I---AN- 143  
Db 840 SCPA-GERCSVQNLGCGYDPRFGTCQSGDPHVSFDGRRRPFDMGTCTCTYLLVSGCCQNA 898  
QY 144 -----N-H-----T-----DMME----- 150  
Db 899 ALPAFVRLVENEHRGSGTQVSYTRAVRVEARGVKVAREYPGQVLDVDDLYLPFQAADG 958

QY 151 -V-D-G-D-V-E-----156  
Db 959 QVQVFRQGRDAVVRTDGLTVTYDMNARVTAKPSSVAEALCGLGNFNGDPADDLALRG 1018  
QY 157 -I-:-----P-P-----N-K-----161  
Db 1019 GGOAANALAFGNSQWQETREPGCATBFGDCFKLDSLVAQLOQSKNCGILADPKPFPREC 1078  
QY 162 -AV--VL-R-----G-H-----168  
Db 1079 HSKLDPOGAVRDCVYDRCLLPGQSGPLCDALATAYAAQAAGATVHPWRSEELCLPSCPP 1138  
QY 169 -----ES-----170  
Db 1139 HSHYEACSYGCLPLSCGDLVPVGGGSECHGCVCEGDFALSGESCLPLASCGVHQGYH 1198  
QY 171 -----E-----V-----172  
Db 1199 PPGQTFYPGDCSLCHQBGGLVSCSSCGPHEACQPSGSLGCVAVGSSTCQASGDP 1258  
QY 173 -F-I-C-----AW-----N-178  
Db 1259 HYTFDGRFRFMGTGVYVLAQTCTGTRPGLHRAVLQENAVANGRVSUTRVTITVQVAF 1318  
QY 179 -PVS--D-L--LAGG--S-G-D--ST--A-----RI--195  
Db 1319 TLRLEQKQWKTVNGVDMKLPVVLANGQIRASQHGSDVVIETDFGLRVAVDLVYVYRVTV 1378  
QY 196 -----W-----196  
Db 1379 PGNYQOMCGLGNYNGDPKDPKPNQSGNAGNANFNGSNWEEVVPDPCLPPTCPGGS 1438  
QY 197 -----NL 198  
Db 1439 EDCIPSHKCPPELEKKYQKEEFCGLLSSPTGPLSSCHKLVDPQGPLKDCIFDCLCGGNL 1498  
QY 199 S-----E--NS-----T-S-G--ST--207  
Db 1499 SILCSNIHAYVSQAAGHVEPWRNETFCPMECPNPNHVELCADTCSLGSALSAPPQC 1558  
QY 208 -Q--L-----VL--R-----212  
Db 1559 QDCABEGCQDSGLYNGQACVPIQCGCYHNGVYEPETQVLIDNCRQOCTCHAGKMW 1618  
QY 213 -H--C-I-----R--E-GGQ-----220  
Db 1619 QBEHCKPGQVQCPSPGGLSCVTKDPCHGVTCRPQETCKEQQGGVCLPNYEATCWLMD 1678  
QY 221 -----D-----V-----P-S-----224  
Db 1679 PHVHFDGRKFDQGTCTNYVLAATTCGPGVSTQGLTPFTVTTKQNRGNPAVSYYRVVVA 1738  
QY 225 -N--KD-----VT-----SLDWN--234  
Db 1739 ALGTNISIIHKDEIGKVRVNGVLTALPVSADGRISVTQAGKALLVADEFLQVSYDNWR 1798  
QY 235 -----S-----E--236  
Db 1799 VDTLPSHYGAVGCLGNMNRPNNDQVFPNGTLAPSIPIWGGSWRAPQWDLPCWDBCR 1858  
QY 237 G--T--L-----LA--TG-----243  
Db 1859 GSCPTCPEDLEQYEGFGPLAPGTGGPFTTCHAHVPSPFPGKCVLDVCMGGGDRDI 1918  
QY 244 -SY-----245  
Db 1919 LKALASYAAQAAGVVIEDMRAQVCEITCPENSHYECGPPCPASCPSAPLATPAV 1978  
QY 246 -----D-GF--A-R-I-----W--T-----K--DGN-----257  
Db 1979 CEGPCVEGQCQDAGFVLSADRCVPLNNGCCWANGTTHAGSEFWADGTCSQWCRCPGG 2038  
QY 258 --L---AS-----TL--GQHKG--PIF-ALK--W-----N-----275

Db 2039 GSVLCTPASCGLEVCGLLPSGQH--GCQPVSTA--EQAMGDPHVYVTLDGHRFPQGTCEY 2096  
QY 276 -----KK-G--NF-----I-----LSA-----GV 286  
Db 2097 LLSAPCHGPGPLGAENFTVTVAHEHRGSAQVSYTRSVTLQIYNHSLTLSARWPKQLQVDGV 2156  
QY 287 -----DK-----TT-----290  
Db 2157 FVTLPFQLDSLHLHAHLSGADVVTITTSGLSLAPDGSFVRLRVPAAYAGSLCGLCGNYNQ 2216  
QY 291 -----II-----292  
Db 2217 DPADDLKAVGKPAQWQVGAQCGSCVSKPCPSCTPEQGESFGGPDACGVISATDGPL 2276  
QY 293 -----W--D-----294  
Db 2277 APCHGLVPPAQYFQGGCLLDACQVQHPGGLCPAVATYVAACQAQAQLREWRPDPFPQ 2336  
QY 295 --AHT-----G-E-A--K-----Q-----302  
Db 2337 CPAHSHVELCGDSCPGSCPSLSAPEGESACREGCVCDAGFVLSGDTCPVPGCGCLHDD 2396  
QY 303 -Q-F-----304  
Db 2397 RYYPQGTFTYPPGCDLRCRREGVSCPSGPHETCRPSGSLGCVAVGSTTCQAS 2456  
QY 305 -P-----F-----H-----S-----308  
Db 2457 GDPHYTFDGRFRFMGTGVYVLAQTCTGTRPGLHRAVLQENAVANGRVSUTRVTITVQV 2516  
QY 309 A-----P-----ALD-----313  
Db 2517 ANFTLRLEQKQWKTVNGVDMKLPVVLANGQIRASQHGSDVVIETDFGLRVAVDLVYVYR 2576  
QY 314 -V-----D-----W-----316  
Db 2577 VTVPGNYQLMCGLCGNYNGDPKDPKPNQSGNAGNANFNGSNWEEVVPDPCLPPTCP 2636  
QY 317 -----Q-----317  
Db 2637 PGSEGCIPSEBCEPPELEKKYQKEEFCGLLSSPTGPLSSCHKLVDPQGPLKDCIFDCLCG 2696  
QY 318 -SN-----N-TF-----A--S--CS--T--327  
Db 2697 GNLSILCSNIHAYVSQAAGHVEPWRNETFCPMECPNPNHVELCADTCSLGSALSAP 2756  
QY 328 -----D-----MC--I-----H-----V-----C--K 335  
Db 2757 LQCPDGCABEGCQDSGLYNGQACVPIQCGCYHNGVYEPETQVLIDNCRQOCTCHAGK 2816  
QY 336 L-----CQ-----D-----R-----I-----342  
Db 2817 VVVCBHSCKPGQVQCPSPGGLSCVTKDPCHGVTCRPQETCKEQQGGVCLPNYEATCWL 2876  
QY 343 -----K-----T-F-----Q-----346  
Db 2877 WGDPHVHFDGRKFDQGTCTNYVLAATTCGPGVSTQGLTPFTVTTKQNRGNPAVSYYRVV 2936  
QY 347 -G-----HTNE--VN-----A--I-----K-----356  
Db 2937 TAAALGTNISIIHKDEIGKVRVNGVLTALPVSADGRISVAQAGKALLVADFLQVSYDW 2996  
QY 357 -W--D--PT-----GNL-----LA-----S-----366  
Db 2997 NNRVDVTLPSHYGAVGCLGNMNRPNNDQVFPNGTLAPSIPIWGGSWRAPQWDLPCW 3056  
QY 367 -C--S-----D-----D--M-----371  
Db 3057 ECRGSCPTCPEDLEQYEGFGPLAPGTGGPFTTCHAHVPSPFPGKCVLDVCMGGG 3116  
QY 372 --TL--K-----I--W-----SM-----378

Db 3117 HDILCKALASYAACAAAGVVIEDWRAQVGCITCPCNSHYEVGPPCPASCPSAPLTT 3176  
QY 379 -----K-Q-----D-N-C-----VH-----D----- 386  
Db 3177 PAVCEGBCVGCQCDAGFVLSADRCVPLNCGCWANGTYHEAGSEFWADGTCQWCRCG 3236  
QY 387 -----L-----Q-----Q----- 389  
Db 3237 PGGSLVCTPASCGLGEVGLLPSGHCQCPVSTAECAQWGDPHYVTLGDHREDFQTCCE 3296  
QY 390 -----H-----N-----KE-----I-YT-----I-----KW----- 399  
Db 3297 YLLSAPCHGPPLGAENFTVTVANERHGSQAVSYTRSTVTLQIYNHSLTLSARWPKLQVDG 3356  
QY 400 -----S-P-----T-----G-----P-----GT-----N-N 408  
Db 3357 VFVTLPLQLSLHLAHLGADVVVTTTSGLSLAPDGSFVRLRVPAAYAGSLGCLGCNYN 3416  
QY 409 --PNA-NLM-----L-----A-----SAS-- 419  
Db 3417 QDP-ADDLKAVGKPKAGVGVGAQCGECVSKPCPSCTPEQOESFGPDACGVISATDG 3475  
QY 420 -----F-----D-----ST-V-----RLM----- 427  
Db 3476 PLAPCHGLVPPAQVYFOGCLLDACQVGHGCLCPAVATYVAACQAAGALREWRPDPFCP 3535  
QY 428 -----P-----D-----V----- 429  
Db 3536 FQCPAHSHYELCGDSFCGSCPSLSAPGCSRSACREGCVCDAGFVLSGDTCPVPGQCCLH 3595  
QY 430 -DR-----G-I-----CI----- 435  
Db 3596 DDRIYPLGQTFYPGPGCDLSLRCRGEVSCPESSCGPHETCRPSGSGSLGVAVGSTTCQ 3655  
QY 436 -----H-----T-----L-----TK-----H-----QE----- 443  
Db 3656 ASGDPHTYTFDGRFDFMGTCVYVLAQTCGTRPGLHRFAVLQENVANGRVSVTRVITV 3715  
QY 444 -----PV----- 446  
Db 3716 QVANFTLRSLRQKWTVGVDMKLPVILANGQIRASQSGSDVVIETDFGLRVAYDLVY 3775  
QY 447 --SV-----A-F-S-----PD----- 453  
Db 3776 VRVTPGNYQLMCLGNYNGDPKDFQKPGSQAGNANEFGNWEVVPDSCPLPPT 3835  
QY 454 -----G-----RY-----LAS-G--S-----PD----- 463  
Db 3836 CPPGSEGCIPSECPPELEKKYQKEEFCGLLSPTGPLSSCHKLVDPQGPLKDCIFDLCL 3895  
QY 464 -----K-C-----V-----HI-W-N-T-----Q-----T--G--AL-- 476  
Db 3896 GGGNLSILCSNIHAYVSACQAAGHVBPWRNETFCPMECPQNSHYELCAUTCSLGCSALS 3955  
QY 477 -----Y-----HS--Y-----R----- 481  
Db 3956 APLOCPDCAEGCQCDGFLYNGACVPIQQCGYHGVYVEPEQTIVLIDNCRCQCTCHV 4015  
QY 482 -----G-----TGGLF-----E-----VC----- 490  
Db 4016 KGVVCOEHSCKPGQVQCPGSGILSCVNDKDPCHGVTCRPOETCKEQGGQGVCLPNYEATC 4075  
QY 491 --W-----N----- 492  
Db 4076 WLKGDPHYSFDRKDFQGTQCNVTLATTCGPGVSTQGLTPFTVTYTKNQRGNPAVSVR 4135  
QY 493 -----AA-G-----D-KV-----CAS-----A-----S- 503  
Db 4136 VVTVAALGTNISHKDEIGKVRVNGVLTALPVSVADGRISVAQAGSKALLVADFGVQSVY 4195  
QY 504 -----D-----GVC-----V-----L-----D-L- 512  
Db 4196 DWNVRVDVTLPSYHGAVCGLCGNMDRNPNDQVFPNGTLAPISIPWGGSWRAPGWDPLC 4255

QY 513 ----R-----K 514  
Db 4256 WDECRGSCPTCPE 4268

## RESULT 10

US-09-334-220-1  
; Sequence 1, Application US/09334220  
; Patent No. 6323177  
; GENERAL INFORMATION:  
; APPLICANT: St. Jude Children's Research Hospital  
; APPLICANT: Curran, Thomas  
; APPLICANT: D'Arcangelo, Gabriella  
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
; TITLE OF INVENTION: THERAPIES  
; FILE REFERENCE: 2427/0F704  
; CURRENT FILING DATE: 1999-06-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 3460  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-334-220-1

Query Match 69.6%; Score 2545.5; DB 3; Length 3460;  
Best Local Similarity 11.5%; Pred. No. 2.7e-32;  
Matches 393; Conservative 90; Mismatches 16; Indels 2915; Gaps 362;

QY 1 M-S-----I-----S-----S-----D-----EV----- 8  
Db 1 MERSGWAROTFLALLLGLATLRARAAGYVPRPFFFLCTHHGELEGDEGEQEVILSLH 60  
QY 9 --N-----F--LV--VRY--LQ--ES-----GF--S-H- 23  
Db 61 IAGNPTVYVPGQEHVYVITSTSTFFDGLLVTLG-LYTSVQASQSIGSSAFGFGIMSDHQ 119  
QY 24 -----S-A-----P-----T-----P-----G--I-K----- 31  
Db 120 FGNQFMCVVASHVSHLPTTNLSFIWIAPPAGTCGVNFMATATHRGQVIFKDALAQOLCE 179  
QY 32 -----S-----H-----I-----S--Q--SNI-----N-----CAL----- 43  
Db 180 QGAPTVTVPHLAEIHSDSIILRDDFDYHQLQLPNIWVECNCEGEGCQCAIMHGNA 239  
QY 44 V-----P--P-----A-AL--I-----SI-----IQ-- 54  
Db 240 VTFCEPYGPRELITTTGLNTTASVLQFSIGSGSCRFSDPSIIVLYAKNNSADWQLEK 299  
QY 55 -----KG--LQY--V-BA-E-----VSIN----- 67  
Db 300 IRAPSNVSTIIHLYLPEDAKGENVOFQWKOENLRVGEVYEAACWALDNILI-INSARQV 358  
QY 68 --ED-----GT-LP-----DG-----R----- 76  
Db 359 VLEDSLDPVDTGNWLPFGATVKHSCQSDGNSIYFHGNEGSEFPNATTDDVLDSTEDIQE 418  
QY 77 -----P-----IES-LSLI-----DA----- 86  
Db 419 QWSEEFESQPTGWDVLGAVIGTECGTIESGLSNVFLKDGKRLCTPSMDTTGYGNLRFYF 478  
QY 87 VM-----P-----D-----D-----VW-----QT----- 94  
Db 479 VMGGIDCPGNSHENDIILYAKIEGRKEHITLDTLSYSYKVPFSLVSVVINPELQTPATKF 538  
QY 95 --RQ--QAY-R--D-----K-LAQ-----O----- 106  
Db 539 CLRQKHQGHNRNVAADFHVLPVLPSTMHMIQPSINLGCCTHOPGNSVLSFSTNHG 598  
QY 107 QA-----A-----AAA-----AA----- 114





QY 422 -S-----T-----VR-----L-----W-D-----VD-RG-----432  
Db 2807 PSVFFPTKGWKRTYPLPESLVGNVRFYQYSDMQMAIDNFYLGPGCLDNCRGHDC 2866  
QY 433 -----IC-----I-HTL-T--KHQ--E--P-----V-----445  
Db 2867 LREQICDPCYSGPNCVLTHTLTKFLKERDSEIKPDLWMSLEGSTCTECGILAEFTA 2926  
QY 446 -Y--S-----V-----A-F-----S--P-----DG-----454  
Db 2927 LYFGGSVRAQVQDLRLCAKELQYWGRISSNNMTSCHRPICKEGVLLDYSTDGGIT 2986  
QY 455 -----R-Y-L--A-----S-----SG-----460  
Db 2987 WTLLEMDYQKYSVRHDYILLPEDALTNTTLRWMPQFVISNGIVVSGVERAQWALDNT 3046  
QY 461 -----S-----K-----C-----VHI-W-N--T---471  
Db 3047 LIGGAENPSQLVDTFDEGTSHENWSFYFNAVRTAGFCGNFPHLYWPNKKDKTHNA 3106  
QY 472 -----Q-----T-G-----A-----LV-----477  
Db 3107 LSSRELIIQGYMQFKIVVGCATSCGLHSLVMLEYTKDARSDSWOLVOTQCLPSSNS 3166  
QY 478 -----H-S-Y-----R-----G--T-----483  
Db 3167 IGCSPFQHEATTYNSVNSSWKRTIQLPDHVSSSATQFRWIOKGEETEKSQSWAIDHY 3226  
QY 484 -----G-----I--PE--VC-----W-N-492  
Db 3227 IGACPKLCSGHGYCTGAICIDESFGDDCSVFSHDLPSYIKDNFESARVTEANWETI 3286  
QY 493 -----A-A-GDKV--G-----ASAS--D-----GS--V--C 508  
Db 3287 QGGVIGSGCQALAPYAHGDSLYFNGCOIROA-ATKPLDLTRASKIMFVLIGMSQTDSC 3345  
QY 509 -----VL-----D-----L--R-K 514  
Db 3346 NSDLSPGHAVDKAVLLQYVSNNGITWHVIAHQHPKDFTOQRVSNVPLEARMK 3399

## RESULT 11

US-09-911-842A-2  
; Sequence 2, Application US/09911842A  
; Patent No. 6656707  
; GENERAL INFORMATION:  
; APPLICANT: Amgen Inc.  
; TITLE OF INVENTION: C3B/C4B COMPLEMENT RECEPTOR-LIKE MOLECULES AND USES THEREOF  
; FILE REFERENCE: 01017/37592  
; CURRENT APPLICATION NUMBER: US/09/911,842A  
; CURRENT FILING DATE: 2001-07-24  
; PRIOR APPLICATION NUMBER: US 60/222,438  
; PRIOR FILING DATE: 2000-08-01  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 2  
; LENGTH: 3571  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-911-842A-2

Query Match 69.5%; Score 2541.9; DB 4; Length 3571;  
Best Local Similarity 11.34; Pred. No. 3.8e-32;  
Matches 403; Conservative 77; Mismatches 18; Indels 3061; Gaps 352;

QY 1 M-----SI-----S-----S-----5  
Db 24 MSPSRNFSPLFPETAPGACGSIAPPAPGDEAGSRVERLGOAFRRVRLLRELSELE 83  
QY 6 -----D-----EYVF-----L-----V-----Y--R--Y-----15  
Db 84 LVFLVDDSSSVGEVNFSELMFVKLLSDFPVPTATRAVAIVTFSSKNYVVRVDYISTR 143

QY 16 -----LOE-----S-----G-F--S--H-----S-----24  
Db 144 RARQHKCALLQEIIPAISYRGGTYTKGAFQAAAILHARENSTKVVFLLITDGYSNNGD 203  
QY 25 -----A-----FTFGI-----K-----S-----H--33  
Db 204 PRPIAASLRDSGVEIFTFGIWQGNIRELNDMASTPKEEHCYLLHSPFEALARRALHED 263  
QY 34 -----I-----S-----Q-----36  
Db 264 LPSGSIQDDMVHCSYLCDEKDCDRMGSCRGTHTGHFECICEKGYGKGLQYECTAC 323  
QY 37 -----SNI--N-----G-A--LV-----P 45  
Db 324 PSPTYKPEGSPGSISSCIPDENHTSPPGSTSPEDCVCREGYRAGSQTCVELVHCFAALXP 383  
QY 46 P-----AA-----LI--SI-----Q 54  
Db 384 PENGYFIQNTCNNHFNACGVRCHPGFDLVGSSIILCLPNGLWSGSESYCRVRTCPLHRQ 443  
QY 55 -----K-----G--L-----57  
Db 444 PKGHISCSCTREMLYKTTCLVACDEGRLEGSDKLTTCQNSQWMDPEPRCVERHCFTOM 503  
QY 58 -----Q-----YV-----60  
Db 504 PKDIIISPHNCGQPAKFGTICVVSRCQGFILSGVKEMLRCTTSGKWNVGVQAACVCKVE 563  
QY 61 -----EA-----E-VS-----I--66  
Db 564 APQINCPKDIEAKTLEQODSANVTWQIPTAKDNGSEKSVHVHPAFTPPYLPFGVDAIV 623  
QY 67 -----N-----D-G-R--P-----78  
Db 624 YTATDLSGNQASCIHFKIVDAEPVIDWCRSPPPVQVSEKVAASWDEPQFSNNGAEL 683  
QY 71 -----T--LF-----D-G-R--P-----78  
Db 684 VITRSHQDGLPQGETIVQYATDPSGNRCDIHIIVKSGCELPFPVNGDFICTPD 743  
QY 79 -----E-----S-----80  
Db 744 NTGVNCTLTLEGYDFTEGSTDKYCYAYEDGVMKPTTYTTEWPCAKKRFANHGFKSPMF 803  
QY 81 -----L--S-----L-----ID-----85  
Db 804 YKAARCDTDLMKKFSFAFETTLGKVPSPFCSDAEDIDCRLEENLTKKYCLEYNYDYENG 863  
QY 86 -AVMP-----DVVQ-----T--94  
Db 864 FAIGPGWGAANLDYSYDDFLDTVQETATISGNAKSSRIKRSAPLSDYKILFINITAS 923  
QY 95 -----R-----QQ-----97  
Db 924 VPLPDERNDTLEWENQORLLQTLTITNKLRKLNKDPMSYFQASEILLIADNSLETCK 983  
QY 98 A-----Y-----R-----D-----KL-----103  
Db 984 ASPFCRPGSVLRGRMCMVNCPLGTYNLEHFTCSGRIGSYQDEEGQLECKLCPSGMYTEY 1043  
QY 104 -----AQ--Q-----Q-----A--108  
Db 1044 IHSRNLSDCKAQCKQGTYSYSGLETCECPGLGYQPKFGSRSCLSCPENTSTVKRGAVNI 1103  
QY 109 AA-----A-A--A-----A-----114  
Db 1104 SAGCVPCPEGKFSRSLGMPCHPCPRDYOPNAGKAFCLACFPYGTTFPAGSRSITECSSF 1163  
QY 115 -----AAA-----AS-----QQ-----G-122  
Db 1164 SSTFSAABESVVPASLGHIKKHEITSSQVHFECFFNCPNCHNSGTCCQLGRGYVCLCPGY 1223

123 SA-K-----NG-----EN-----129  
1224 TGLKCTDIDECSPCLNNGVCKDLVGEFICEPCSGYTQORCBENINECSSPCLNKGI 1283  
130 -----T-----AN-----G 133  
1284 CVDGAGYRCTVKFVGLHCETEVNECOSNPLNNAVCBQVGGFLCKCPGFLGTRCG 1343  
134 -----E-----ENGA-----H-----TI-----A-----N 143  
1344 KNVDECLSQCKNGATCKDAGNSFRCLCAAGFTGSHCELNINECOSNPCRNOATCVDEN 1403  
144 -----N-----N-----H-----T-----DM 148  
1404 SYSCKQPGSGKRCETEQTGFNLDFEVSIGYVMDGLPSLHALTCTFWMKSSDDM 1463  
149 -----M-----E-----V-DG-----153  
1464 NYGTPISYAVDNGSDNTLLLTIDYNGWLYVNGREKITNCPNSVNDGRWHIAITWTSANGI 1523  
154 -----D-----V-----E-----I-----157  
1524 WKYVIDKLSGGAGLSVGLPIPGGALVLQGDQKKGEGFSPAESFVGSISQNLNWDYV 1583  
158 --P-----P-----N-----KA-V-----VL-----R-G--H-- 168  
1584 LSPQOVKSLATSCPEELSKGNVLAWPDLFSLGIVGKVIDSKSIFCSDCPLGGSVPHLRT 1643  
169 --E-----SEV-----P-----I-C-----AWN--P-----V-----S 181  
1644 ASEDLKPGSKVNLFCDFGOLFQVGNVQYCLNOGQWTPQLPCHBERISCGVPPPLENGFHA 1703  
182 -DLASGS-----GDSTARI-----WN-----L-----SE-- 200  
1704 DDFYA-GSTVYQCNNGYVLLGDS--RMFCTDNGSNVNGVSPCLDVEDCAVGDCEHAS 1760  
201 --N-----S-----T-----SG--S-----T-----Q 208  
1761 CLNVDSYICSVPPVTGDKNCAEPIKCKAPGNENGSHSGEITYVGAETVSCQEGYQ 1820  
209 L-V-----L-----R-H-----CI-----R-EG-- 218  
1821 LMGVTKITLESGEWNHLIPYKAVSCGKPAIPENGCIELAFITFGSKVYTRCNKGYTLA 1880  
219 GQB-----V--P--S-----N-KDV-----T-SL--D-----232  
1881 G-DKESCLANSWSHSPVCEPVKCSSPENINNGKVIISGLTYLSTASYCDTGYSLOG 1939  
233 -----W-----NS-----EG-TL-----239  
1940 PSIICTASGIWDRAAPPACHLVFCGBPPAIPKDAVITGNFTFRNTVYTKCEGYTLAGLD 1999  
240 -----LA-----T-----G-----S-Y-----245  
2000 TIECLADGKWSRSDQCLAVSCDEPPIDVHASPETAHRLFGDIAFYCYSDGYSLADNSQL 2059  
246 -----D-----GF-----248  
2060 LMAQCKWVPPBEGQDMPRCIAHCEKPPSPSVSISLESVKAKFAAGSVSVFKMGEFVLN 2119  
249 --ARI-----W-----T-KDG-----NL-----A-----259  
2120 TSAKIECMRGQWNPSPMSITQICPVRCGBPPSIMNGVASGNSYVFCGAWAYSCKNGFYIK 2179  
260 --ST--LQO-----HK--G--P-----IF-----ALK-- 273  
2180 GEKSTCEATQWSSPIPTCHPVSCGEPKPVENGFLHTTGRIFSEVRYQCNPGYKSVG 2239  
274 -----W-----N--KKG-NP-ILS-----AG--V-D 287  
2240 SPVFVCOANRHHSESPLMCVPLDCGKPPPIQNGFMKGFENFVGSKVQFFCNSGELVGD 2299  
288 -----K--TT-----II-----292

Db 2300 SSWTCQSGKWNKSNPKCMKPAKCPBPPLLENQLVLKELTTEVGVVTFSCKEGHVLQGPS 2359  
QY 293 -----W-D-----A-H-----296  
Db 2360 VLKCLPSQQMNSFPVCKIVLCTPPPLISFGVPPISSALHFGSTVKYKSCVGGFLLRGNST 2419  
QY 297 T-----G-----E-AK--Q-----Q-----F--P--F-----306  
Db 2420 TLCQPDGTWSSPLPECVPECPQPEIPNGIIDVOGLAYLSTALYTCCKPFELVGNNTTL 2479  
QY 307 --H-----S--A-----P-----AL-D-----V-----314  
Db 2480 CGENGHWLGKPTCKAIECLKPKRIILNGKFSYTDLHVGQTVTYSCNRGPRLEGPSALTCL 2539  
QY 315 --DW--QS-N-----N-----T-----F--A--SC-- 325  
Db 2540 ETGDMVDVAPSCNAIHCDSPQPIENGFEVADYSYGAIYYSCFPGFQVAGHAMQTCEES 2599  
QY 326 --S--T-----DMC-I--HV-----C-KL-----G--Q--D-----339  
Db 2600 GWSSSIPTCMPID-CGLPHIDFGDCTKLKDDQGYFEQEDDMMEVPVYTPHPVHLGAVA 2658  
QY 340 -----F-----Q-----GH--T-----N-----E-----351  
Db 2719 LPTAPENGFLRTETSMGSAVQSKPGHILVGSDLRLCLLENKRWKSGASPRCEAISCKKP 2778  
QY 352 --V-N-AIK-----N-----LL--A-S--C-----S--D-----369  
Db 2779 NPWNGSIKGSNYTYLTYECDPGYVLNGTERRTCQDDKNWDEDEPICIPVDCSSPPV 2838  
QY 360 T-----G-----N--LL--A-S--C-----S--D-----369  
Db 2839 SANGVRGDEYTFQKEIETCTNEGFLLEGARSVCLANGWSGATPCVPRCATPQLA 2898  
QY 370 -----D--W--T-----L-----KI-----WSMKOD-----NC-----383  
Db 2899 NGVTEGLDYGFKEVTFHCHEGYLHGAPKLTQCSQGNW-----DABIPCKPVNCGPPED 2954  
QY 384 -----VHD--LQ-Q-----H-----N-----KE--I-Y 395  
Db 2955 LAHGPNGFSFINGHIOYQCPFGYKLGNSRRCLNSGWSGSSPSCULPCRSTPVEY 3014  
QY 396 --T-----I-----K-----WS--P-----T--G--PGTNPN 410  
Db 3015 GTVNGTDFDCGKAARIQCFKGLLSEITCEADQWSSGSPHCBHTSCGSLP--MIPN 3072  
QY 411 A-----N-LMLA--S-----ASPD-----S-----422  
Db 3073 AFISSETSKWKNVITYSCRSYVIOGSS-DLICTEKGWNSQPYVPVCEPLSCGSPPSVANA 3131  
QY 423 -----T-----V--R-L--W--DVD--R-G-----I--C-----I-435  
Db 3132 VATGEANTYSEVKLECYMTDITDFTCKDGRWPPERISCSKPKPLPENTHIL 3191  
QY 436 --H-----T-----L--T-----K-----H-- 441  
Db 3192 VHGDSPVNRQSVSCAEGYTFEGVNIISVCQLDGTWEPFSDSCSPVSCGKPEHGF 3251  
QY 442 -----Q-EP-----V-----YS--VAP-----SP-----D 453  
Db 3252 VVGSKYTFESTIYYQCEPGYELEGNRVRVCOENROWSGGVAICKETRCETPFLNGKAD 3311  
QY 454 -----G-----R-Y-L-----A--S--G-----S-461  
Db 3312 IENRTTGNVTVSCNRGYSLEGSEAHCTENGTHWHPVLPCKPNPCVPFVIPENALLSE 3371  
QY 462 --F--D-----KC-----V--H--I-----WNTOT-----G-----AL--V 477

Db 3372 KEFYVDQNSIKREGFLQGHGIIITCPDET-TOTSACEKISCPPAHVENAIARGV 3430  
Qy 478 HSYR-G--T-----G--GIF--EVC-----W-----NA- 493  
Db 3431 H-YOYGMITVSCYSGWLEG-FLRSVLENGWTSPPIRAVCRFPCCQNGGICQRPNAC 3488  
Qy 494 -----AG--DK--V-----G--ASAS--D-----GS-----V-----CV 509  
Db 3489 SCPBGWGRLECEPICILPCLNGRCV-APVQCDPCPGWTGSRCHTAVCQSPCLNGGKCV 3547  
Qy 510 -----L-----DL-RK 514  
Db 3548 RPNRCHCLSSWGTGNCGRK 3566

## RESULT 12

US-08-471-119A-2  
; Sequence 2, Application US/08471119A  
; Patent No. 5827706  
; GENERAL INFORMATION:  
; APPLICANT: Leitner, Ernst  
; APPLICANT: Schneider, Elisabeth  
; APPLICANT: Schoergendorfer, Kurt  
; APPLICANT: Weber, Gerhard  
; TITLE OF INVENTION: Cyclosporin Synthetase  
; NUMBER OF SEQUENCES: 8  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: No. 5827706artis Corporation  
; STREET: 59 Route 10  
; CITY: East Hanover  
; STATE: New Jersey  
; COUNTRY: USA  
; ZIP: 07936  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/471,119A  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kasenoff, Melvyn  
; REGISTRATION NUMBER: 26,389  
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT  
; TELEPHONE: 201 503 8474  
; TELEFAX: 201 503 8807  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 15281 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Tolypocladium niveum  
; STRAIN: ATCC 34921  
US-08-471-119A-2

Query Match 69.4%; Score 2539.8; DB 2; Length 15281;

Best local Similarity 7.6%; Pred. No. 1.1e-28;  
Matches 469; Conservative 39; Mismatches 3; Indels 5662; Gaps 374;

Qy 1 M-----S-----ISSDE-----VNF-----L-----V 12  
Db 5261 MKATNELSSRYAAVLHI-SDEPLPIYKIDPEAWINFEGLRLEALAQVLKENAESV 5319  
Qy 13 -----Y-----R-----YL-----Q----- 17

Db 5320 AISNIPIYKTVVERHIVRSIDQEDANAPESMDGSDWISAVRTAQOCHTSLASDLFDIA 5379  
Qy 18 E-SGF--S-----HS-----A-F-----T-----F----- 28  
Db 5380 EDAGFRVEVSWARQSHQSGALDAVFFHLKPATEDSRVLKFPDTHQGRPLKSLTNQPLLP 5439  
Qy 29 -----G-----G-----G-----I-KS----- 32  
Db 5440 AQSRRAELLIREGLQTLPPYMIPTSLIDRMLNANGKVDREELARRAKITQKSPVE 5499  
Qy 33 -----H-----H-----H-----IS----- 35  
Db 5500 DIVPPRNSVEATVCKGFTDVLGVEVGITDNFNLGCHSLMATKLAARLGRQLNTRISVRD 5559  
Qy 36 -----QSN-----I-----N-GA----- 42  
Db 5560 VFDQPVVADLAAVIORSAPHEPIKPADYTGVPVQSFQAQRLWFLDQLNVGATWYLMPLG 5619  
Qy 43 -----L-----V-----P-----P----- 46  
Db 5620 IRLHGLRVDALATAISALEQRHEPLRTTFHEEDGVGVVQVQDHRPKDLRIIDLSTQPKD 5679  
Qy 47 -----A-----ALI-----S-----I----- 53  
Db 5680 AYLAVLKHEQTTLFDLATEPGMRVALIRLGEHEHILSIVMHHIISDGSVEVLFDGMHRF 5739  
Qy 54 -----Q--K--GL--QY-----V-----E-----AEV----- 64  
Db 5740 YSSALRQDPMEOILPFIQYRDFAAWQKTEQVAEHQRLQDYWTEHLADSTFAELLTDL 5799  
Qy 65 -----SI-----N-----L-----F-----D----- 70  
Db 5800 PRESIILSGRANELPLTIEGLHDKLRAFCEVHQATPFVILLAAALRAAHYELTCAEDATLG 5859  
Qy 71 T-----N-----L-----F-----D----- 74  
Db 5860 TPIANRNPENLENMIGFFVNTQCMRIAIBENDNFESLVRVRSTATSAFANQDVPFESIV 5919  
Qy 75 -----G-----R-P-----I-----E-----S----- 80  
Db 5920 SLLPGSRDASRNPLVQVILAVHSQQDLGKLTLEGLRDEAVDSAISTRFDVPHLFEHAD 5979  
Qy 81 -LS-----L-----I-----D-----AVMP-----DV 91  
Db 5980 RLSGSVLYAKELFKLRTIESVSVFLETARRALDQPLTPLAVLPLDTGVGEIASKGLDV 6039  
Qy 92 -----V--Q--TR-----QO----- 97  
Db 6040 PRTDYPDANIVEVFOQHVRAATPDALAVKDATSILTVAQLDQOSDRLAIWLSRRHMPET 6099  
Qy 98 -----AV-----R-----D-----KL-----A- 104  
Db 6100 LVGLAPRSCETIIMFGIMKANLAYLPLDINSIPAARLSILSAVDGNKLVLGSGVTAP 6159  
Qy 105 -Q-----Q-----Q-----A-----A----- 110  
Db 6160 EQENPEVEAVGIEIILAGTGLDKTQCSNAPSPATSLAYVIFTSGSTGPKGVNVEHRSVT 6219  
Qy 111 --A-----A--A--A--A--A-----A----- 118  
Db 6220 RLAKPSNVISKLPQGARVAHLANIAFDASIWETATLLNGATLVCLDYHTVLDCTLKEV 6279  
Qy 119 -----S-----Q-----Q----- 121  
Db 6280 FERESTVVTMLPALKQCVAEIPETLAHLDLLYTGDRVGGHDAMRAKSLVKIGMPSGY 6339  
Qy 122 G-----S-----A-----K-----N-----GE----- 128  
Db 6340 GPTENTVISTIVEVDADMFVNGVPIGKTVNSGAYVMDRNOQLVPSGVVGVGELVVTGDL 6399  
Qy 129 -----N-----TANGE----- 134  
Db 6400 ARGYDPSLNKRFIYITVNGESIRAYRTGDRVVRPHDLQIEFFGGRMDQOVKIRHRIE 6459

QY 135 E--N--GA-- 138  
Db 6460 PGEVSAALLSHNSVDAAVVICAPADQSGAEMVAFVAARNTEDTQBEAAVDQVQGW 6519  
QY 139 H--T--I--ANN--H--T--D-- 147  
Db 6520 THPETAAYSEVKDIROSEVGNDFMGWTSYDGSIDKTDHFWLNDTMRMILDAREPGHV 6579  
QY 148 M--M--E--V--DG--D 154  
Db 6580 LEIGTGTGMWFMENAKCPGLQGVGFEPKSAQFVNDAAQSPALKDGRSIVHVGTAID 6639  
QY 155 INKAGPIQPRLVINSVAQYFPTPEYLFRVVEALVQIPSVRIVFGDMRTNAINRDFVAS 6699  
Db 6640 INKAGPIQPRLVINSVAQYFPTPEYLFRVVEALVQIPSVRIVFGDMRTNAINRDFVAS 6699  
QY 159 P-- 159  
Db 6700 RALHTLGEKANKRLVRQMIYELBANBEELTDPAFTSLRTRIGEIKHVEILLPTMKAT 6759  
QY 160 N--K--AVL--RG--H--E-- 169  
Db 6760 NELSKRYAAVLHVRGSRQSTHQVSPNAWIDFAADGLDRQTLINLLKEHKDAGTVAIG 6819  
QY 170 SE--V--FI--C-- 175  
Db 6820 NIPYSKTIIVERFVNSLSDEDDMEGQSLDGSAAVRAVMAAQCSPSLDAMDVKEIAQEA 6879  
QY 176 A--W--N--P--V--SD-- 182  
Db 6880 GYQVEVSWARQWONGALDAIFHHFEPKPEGARTLIEFPTDYEGRNVTILNRPLNSIQS 6939  
QY 183 LL--A--S-- 186  
Db 6940 RRLGTQIREKLOTLPPYIPMIMVLDQMPVNNNGKIDRKELVRAIVAPKPRSAATRV 6999  
QY 187 G--S--G-- 189  
Db 7000 APRNEIEAILRDEFEDVLGTEVSLDNFDFDLGHSLSMATKLAARVSRRLDAHISIKDVED 7059  
QY 190 D--S--T--A--RIW--NL-- 198  
Db 7060 QPVLADLAASIORESAPHEPIQORPYTGPAEQFAGRLWFLDQNLGATWYLMPLAIRI 7119  
QY 199 198  
Db 7120 RGQLRVAALSAAULFALERRHETLRTTTFESDGVGVQVIGEARNSDLRVHVDVSTGDDGEVL 7179  
QY 199 SE-- 200  
Db 7180 EVLRREQTVFPDLSSBPGWVRVCLVKTGEEDHVLISVMHHIIVDGSVDILRGELGFYSA 7239  
QY 201 N--N--ST-- 203  
Db 7240 ALRGQDPLLHANPLPIQYRDFAAWQREAKQVEEHQRLGWYKQVLDSTPAELLTDLPRP 7299  
QY 204 S--GS--T--Q--L--VLR--H-- 213  
Db 7300 SILSGRAGSYDVTIEGVSYGALQSCFRTSRVTTFFVLLTVFRIAHFRLTAVDATIGTPI 7359  
QY 214 C--I--R--EG-- 218  
Db 7360 ANNRPELETIVGCFVNTQCMRISIADDNDFEGLVQRVNVATAAVANQDVPFERIVSAL 7419  
QY 219 G--Q--D--VP-- 223  
Db 7420 VPGSRNTRNPLVQMFVQSVEDYDQVRLEGLSVMPGCEASTRDMFEHLVPGQKLT 7479  
QY 224 S--N--K--D--V--TS 230  
Db 7480 GSVLYSSDLFEQGTIONFVDFQECRLRSVLDQPLTPIPVLPFSNAISNLSLDLEMPST 7539

QY 231 LD--W-- 233  
Db 7540 DYPRDRTVVDLFRQAAICPDSSIANKDSSQLTYAQDLQSDRVAAMLHERHMPAESLVG 7599  
QY 234 N--S--EG-- 237  
Db 7600 VLSPRSCETIIAYFGIMKANLAYPLDVTYAPDARLAAILDVTVEGERKLLLGAGVQPGIQ 7659  
QY 238 T--L--LA--TG-- 243  
Db 7660 IPRLTAYIAEALSHATTVDVTISIPQPSATSLAYVIFTSGSTCKPKGVMLIEHGRIVLVR 7719  
QY 244 S--Y-- 245  
Db 7720 DTNNVFPBSSGALPVSHFSLNLAWDAATWEIYTAVLNGGTVCIDRDTMLDIAALNSTFR 7779  
QY 246 D-- 246  
Db 7780 KENVRAAFPTAPFLKOCIAETPELVANLEILHTAGRLDPGDANLAGKTAKGGIFNVLGH 7839  
QY 247 G--F--AR 250  
Db 7840 TENTAYSTFPVVGEEFTFVNGVPGRGISNSHAYIIDRHQKLPAGVMGELIILTGDGVAR 7899  
QY 251 I--W--T--K--DG-- 256  
Db 7900 GYTDSALNKDRFVYIDINGKSTWSYRTGDKARYRPRDQGLEFFGRMDQMVKIRGVRIEPG 7959  
QY 257 256  
Db 7960 EVELTLLDHKSVLAATVVVRPPNGDPEMIAFITIDAEDDVQTHKAIYKHLOGILPAYMI 8019  
QY 257 N--LA--ST-- 261  
Db 8020 PSHLVILDQMPVTDNGKVDKRLALRAQTQVKRSTAAARPPRDEVEAVLCEBSYNLLEV 8079  
QY 262 LG--Q--HK-- 266  
Db 8080 EVGITDGFPLGHSLSLATKLAARLSQNLTRVSVKVDQDPIADLADIIRRGSHRHP 8139  
QY 267 GPI--FA--LKW--N--KKG--N-- 279  
Db 8140 IPATPYTGVEQSPAGRL-WFLEQLNGASWYLMFPAIRMRGLOTLKALAVANALVHR 8198  
QY 280 F--I--LS--AGVD--K--TT-- 290  
Db 8199 HEALRTTFEDHGVGVQVIOPKSSQDLRIIDLSA-VDDTAYLAALKREQTTFADLTSEP 8257  
QY 291 II-- 292  
Db 8258 GWRVSLRLGDDDDYILSIVMHHIISDGTVDVLRLQELGFYSAAIRGOEPLSQAQSLPIQ 8317  
QY 293 W--D-- 294  
Db 8318 YRDFAVWQRENQIEQAKQLKYWSQOLADSTPEFLTDLPRPSILSGEADAVPMVIDGT 8377  
QY 295 AH--TG--B-- 299  
Db 8378 VYQLLTDFCRTHQVTSFVLLAAFRTAHYRLTGTLDATVGTPIANRNRPELEGLIGFFVN 8437  
QY 300 A--K--Q--F--PF-- 306  
Db 8438 TQCMRMAISETETESLVQVRLTTTTEAFANQDVPFEQIVSTLLPGSRDTSRNPVQVMF 8497  
QY 307 H--S-- 308  
Db 8498 ALQSQODLGRIOLEGMTDEALETSLRLDLEHVFQEVKLSGSLIYSTDLFEVETIRG 8557  
QY 309 A--PAL--D--V-- 314  
Db 8558 IVDVFLIEILRRGLEQPKQRLMAMPITDGTIKLRDQGLLTVAKPAYPRESSVIDLFRQOVA 8617  
QY 315 D--W--QSN--N--T--FA--SC-- 325

Db 8618 AAPDAIAMDSSSLTYADLDGQSNKLAHMLCQENMAPETLIVAFPRSCLTIVAFGLVL 8677  
Qy 326 -----  
Db 8678 KANLAYPLDVNAPAARIEAIIASVPGHKLVLQAHGPELGLTMADETVELOIDEALASS 8737  
Qy 326 -----  
Db 8738 SGDHEQIHASGPTATSLAYVMTSGTGKPKGVMIDHRSIIRLVKNSDVTATLPTVRMA 8797  
Qy 330 -----  
Db 8798 NVSNLAFDISVQEIYTALLNGTGLVCLDYLTLDSKILYNVFEAQNAAVFPVLLKQC 8857  
Qy 335 -----  
Db 8858 LGNPAIISRLSVLNVGDRDLDAHDAVAASGLIQDAVYNAYGPTENGQMTWKVDVNEP 8917  
Qy 340 -----  
Db 8918 FVNGVPIGRITNSGAVYMDGNQOLVSPGVNGEIVVTGDLGARGYDTSALDEDRFVHTI 8977  
Qy 342 -----  
Db 8978 DGEENIKAYRTGDRVVRPKDFEFIEFFGRMDQVKIRGHRIBPAEVEHALLGHDLVHDA 9037  
Qy 349 -----  
Db 9038 VVLRKANQPEMFIATFITSQEDETIEQHSNKQVQWGEHDFVRSYADIKDLDTSTFGHD 9097  
Qy 357 -----  
Db 9098 FLGWSMYDGVDPVNMKEWLETTASLLDNRPFGHILEIGAGTGMILSNLKGVDGLQK 9157  
Qy 364 -----  
Db 9158 YVGLDPAPSAAI FVNEAVKSLSPLAGKARVLVGTALDIGSLDKNEIOPELVVINSVAQYF 9217  
Qy 366 -----  
Db 9218 PTSEYLIKVVKAVVEVPSVKRVFFGDIRSQALNRDFLAARAVRALGDNASKEQIREKIAE 9277  
Qy 370 -----  
Db 9278 LEESEELLVDPAFFVSLRSLQPNIKHVEVLKMKATNELSSRYAAVLHISHNEEQ 9337  
Qy 375 -----  
Db 9338 LIQIDPTAWVDFPAATQKDSQGLRNLQOGRDDVMIAGVNI PYSKTIVERHIMNSLDQDH 9397  
Qy 382 -----  
Db 9398 VNSLDGTSWISDARSAACITSFDPALATQLAKEEGRVBLSWARORSQNGALDAVPHRL 9457  
Qy 382 -----  
Db 9458 ATDANCERSVLVHPFDHOGRLTLTNRPLOQAQRRIESQVFEALQALPAYMIPSR 9517  
Qy 389 -----  
Db 9518 IIVLPQMPNTNANGKVDRLARRAQVVAKRKAVSARVARNDETIVLCEYADILGTEVG 9577  
Qy 389 -----  
Db 9578 ITDNFDMGHSMATKLAARLSRLDTRVTVKEVPKPLADLAASIEQGSTPHLP IAS 9637  
Qy 389 -----  
Db 9638 SVYSGPVEQSAQRLWFLDQFNLNATWYHMSLAMRLGLPNDALDVALRALEQRHETL 9697  
Qy 390 -----

Db 9698 RTTTEAQKDIGVOVVEAGMKRLKVLDSKNEKEHMAVLENEQMRPFTTLASEPGWKHGL 9757  
Qy 394 -----  
Db 9758 ARLGPTETIYLSLVHMHFSDGNSVDILROELGOFYSAALRGRDPLSQVKPLPQYRDFAA 9817  
Qy 398 -----  
Db 9818 WQKEAAQVAEHERQALAYWENQADSTPGBELLTDPFRPQFLSGKAGVIVPTIEGVPVEKLL 9877  
Qy 405 -----  
Db 9878 KPSKERQVTLFSVLLTAFRATHFELTGAEDATIGTPIANRNRPELEHIIIGFFVNTQCMRL 9937  
Qy 411 -----  
Db 9938 LLTGSTFESLVQHVRSVATDAYSNQDIPPERIVSALLPGSRDASRPLIQLMFALHSQP 9997  
Qy 415 -----  
Db 9998 DLGNITLEGLEHERLPTSVAITRDFMEFHLFOEPNKLSSILFADELFOPETINSVVTVFQ 10057  
Qy 421 -----  
Db 10058 EILRRGLDQPOVSISTMPLTDGLIDLEKGLLEIESSNPRDYVVDVFRQOVAANPNAP 10117  
Qy 426 -----  
Db 10118 AVVDSETSMTSLDQKSEQIAAMHAQGLRPESLICYMAPRSFETIVSLFGILKAGYAY 10177  
Qy 428 -----  
Db 10178 LPLDVNSPAARIQPIILSEVEGKRLVLLGSGIDMPQSDRMDVETARIQDILNTKVERS DP 10237  
Qy 432 -----  
Db 10238 MSRPATSLAYVIFTSGSTRPGKVMIEHRNILLRVKQSNVTSQLPQDLRMAHISNLAFD 10297  
Qy 432 -----  
Db 10298 ASIWEIFTAILNGGALICIDYFTLLDSQALRTTFEKARVNATLAPALLKECLNHAFTLF 10357  
Qy 437 -----  
Db 10358 EDLKVLVYIGGDRLDATAAKIQALVKGVYVNAVPTENTVMSTIYRLTDGESYANGVPIG 10417  
Qy 437 -----  
Db 10418 NAVSSSGAYIMDQKRLVPPGVMGELVWSGDGLARGYTNSTLNADRFVDIVINDOKARAY 10477  
Qy 439 -----  
Db 10478 RTGDRTRYPKGSIIEFFGRMDQOVKIRGHVPEPAEVEQAMLGNKAIHDAVVVQAVDQ 10537  
Qy 445 -----  
Db 10538 ETEMIGFVMSADRFSEGBEEITNQOEWEDHFEFESTAYAGIEAIDQATLGRDFTSMTSY 10597  
Qy 448 -----  
Db 10598 NGNLIDKAEMEELDDTMSQLDKEDARPCARIGTGTGMVLFNLPKNDGLESYVGEIPSR 10657  
Qy 453 -----  
Db 10658 SAALFVDKAAQDPFGLQGTQIILVGTAEADIKLVKDFHPDVVVINSVAQYPPSRSYLVQIA 10717  
Qy 459 -----  
Db 10718 SELIHMTSVKTIFFGDMRSWATNRDFLVSRALYTLGDKATKDOI RQEVARLEENEDLV 10777  
Qy 465 -----  
Db 10778 DPAFFTSLTSQWPGKVKHVEILPKRMRTSNELSSRYAAVLHICRDEGRNRYGRVHSV 10837

QY 469 ---W---NT-----471  
Db 10838 EENAWIDPASSGMDRHVQLMDERRDAKTAIGNIPHSNTINERHFTTSLDTGEGIAQ 10897  
QY 472 ---Q---T-----GAL--V--H--478  
Db 10898 DSLDGSAMQSAKMAARCPCLSVTELVEIGQAAGFRVEVSWARQSORHGALDVFHLE 10957  
QY 479 ---S---S---Y-----480  
Db 10958 DDRVGRVLNFPDPERLPSTGLTSRPLQIONRRFESQIREQLOTLPLPPYMPSRIVV 11017  
QY 481 ---R-----G-T-----483  
Db 11018 LERMLNANSKVDKELARKARTLOTIKPSATRVAPRNDIEAVLCDFQAVLGVTVGMD 11077  
QY 484 ---CG---IP-----E-----488  
Db 11078 NPFELGGHSLMATKLAARLSRLDTRVSVKDIENQPILODLADVQVOTGAPHEAIPSTPY 11137  
QY 489 ---VC---NA---A---G-----495  
Db 11138 SGPEQSFQGRNLWFLDQLNLNASWTHMPLASRLRGLRLEALQSMALATIEARHESLRTT 11197  
QY 496 ---D---K---V-G---A---SA-----502  
Db 11198 FEEQDGVVPQIVRAARNKQLRIIDVSGTEDAYLAALKQEQDAAFDLTAEPGWRVALLRLG 11257  
QY 503 ---SG-SV-----507  
Db 11258 PDDHVLISVMHHIISDGSVDILRLQBLGOLYSNASSQPAPLPIQYRDFAIWQKDSQIAE 11317  
QY 508 ---C-----508  
Db 11318 HOKQLNWKRLVNSKPAELLADFTPRKALSGDADVIPIEDDQVYQNLRSFCRAHVTSS 11377  
QY 509 -V-----L---D-----LR---K 514  
Db 11378 FVALLAAFRAAHYRLGAEDATIGSPIANNRPELSGLICGFVNTQCLRIPVK 11430

RESULT 13

US-08-460-751-2  
; Sequence 2, Application US/08460751  
; Patent No. 5891628  
; GENERAL INFORMATION:  
; APPLICANT: Readers, Stephen  
; APPLICANT: Schneider, Michael  
; APPLICANT: Glucksmann, Sandra  
; TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY  
; TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESS: Pennie & Edmonds  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036-2711  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/460,751  
; FILING DATE: 02-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/413,580  
; FILING DATE: 03-MAR-1995  
; ATTORNEY/AGENT INFORMATION:

; NAME: Coruzzi, Laura A.  
; REGISTRATION NUMBER: 30,742  
; REFERENCE/DOCKET NUMBER: 7638-005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 790-9090  
; TELEFAX: (212) 869-8864/9741  
; TELEX: 66141 PENNIE  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4303 amino acids  
; TYPE: amino acid  
; TOPOLOGY: unknown  
; MOLECULE TYPE: protein  
; US-08-460-751-2

Query Match 69.4%; Score 2539.5; DB 2; Length 4303;  
Best Local Similarity 9.9%; Pred. No. 1.2e-31;  
Matches 414; Conservative 81; Mismatches 14; Indels 3675; Gaps 356;

QY 1 MSIS-----S-----D-----EVN-----F-----10  
Db 72 LDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEGIFANLNFNLSEINLSGNPFECDC 131  
QY 11 -LVY-----R-----Y-----15  
Db 132 GLAWLPQWAEQOVVVQPEAATCAGPGSLAGOPLLGIPLLDSCGEGEYVACLPDNSSGT 191  
QY 16 -----LQ-----E-----SG-20  
Db 192 VAAVFSAAHEGLLOPEACSAFCSTGTQGLAALSEQGWCGLCGAAQSSASFACLSICSGP 251  
QY 21 -----F-S-----H-----S-----AF-----T-----FG-----29  
Db 252 PAPPAPTCHGPTLLQHVFPASPATLVGPHGLASQLAFAHIAAPLVTDTTWDGDS 311  
QY 30 --I-----KSH-----I-----S-QSN-38  
Db 312 AEVDAAGPAASHRYVLPGRYHVTAVALGAGSALLGTDVQVEAAPAALELVCPSSVQSD 371  
QY 39 -----I-N-----G-----AL-----V-----P-----46  
Db 372 SLDLSIQNRGSGGLEAAYSIVALGEPARAVHPLCPESDTEIFPGNGHCYRLVVEKAALQ 431  
QY 47 -----AA-----LIS-----I-----IQ--K-----G-----L--57  
Db 432 AQEQCAWAGALAMVDSPAVORFLVSRVTRSLDWIGFSTVQGVVEGPAQGEAFSLES 491  
QY 58 -Q-----EDG-----YV-E-----AE--V--S-65  
Db 492 CQWMLPGEPHAPATAEHCVRLGPTGMCNTDLCSPHYSVCELOPGGPVQDAENLLVGAPSG 551  
QY 66 -----IN-----EDG-----TL--F-----73  
Db 552 DLQGLTPLAQDGLSAPHEPVEVMVFPGLRLSREALFTAEFGTQELRRPAQLRLQVYR 611  
QY 74 -----D-----GR--P-----I-----78  
Db 612 LLSTAGTPENGSEPSRSPDNRTQLAPACMPGRCWPCGANICLPLDASCHPOCANGCTS 671  
QY 79 -----E-S-----LSLI-----DAV--M--P--DVV-----92  
Db 672 GPGLPGAPVALWREFLFSVPAGPPAQS-VTLHGQD-VLMLPGDLVGLQHDAGPGALLHC 729  
QY 93 -----Q-----T-----R--Q--Q-----97  
Db 730 SPAGHGPAPYLSANASSWLPPLPAQLEGTWGCPACALRLAQRLEQTLVLLGLRPNPG 789  
QY 98 -----A-----Y-----RD-KL-----A--Q-105  
Db 790 LRLPGRYEVAEYGVNGVSRHNLSCSFDVVSPVAGLRYVPAPRDRGLRYPTNGSALVLQV 849  
QY 106 -----QQ-----A-----A--A-----A-----112





QY 381 ---D-N---C---V---H---D-L---Q----- 388  
Db 3068 EPTADVNYVMTATCAVCLVTYMYMAAILHKLQDLDASGRAIPFCQGRFKFYELLVKTG 3127  
QY 389 ----- 388  
Db 3128 WGRSGTTHVGMVIMLYGVDSRSHRHLDDGDFRHNSLDIPRIATPHSLGSLVMKIRVWHD 3187  
QY 389 ---QH--- 393  
Db 3188 NKGLSPAWFLQHVIVRDLQARSAPFLVNDWLSVETEANGGLVEKEVLAASDAALLRFR 3247  
QY 394 ---IY---T-I-K-----W----- 399  
Db 3248 LLVAELQGFDFKHIMLSINWDRPSRFTRIQATCCVLLICLFLGANAVMYGAGDSAY 3307  
QY 400 -----SPT----- 402  
Db 3308 STGHVSRSLPVDVTVAAGLVSSVVPVYLAIFLFRMSRSKVAGSPSPTRAGQQVLDI 3367  
QY 403 -----G-----P-GT-----NNP 409  
Db 3368 DSCLDSSVLDSSFLTSGLHAEQAFVGOMKSDLFLLDCKSLVCPSPGEGTLSWPDLLSDP 3427  
QY 410 ---NANI---M-----LAS---A-SF----- 420  
Db 3428 SIVGSNLRLARGAAGHLGPEEDGFLSAPYSPAKSFASDEDLIQVLAEGVSPAPT 3487  
QY 421 ---D---ST-----V----- 424  
Db 3488 QDTHMETDLLSSISPTGKETTETLQRLGELGPPSPGLNWEQFOAARLSTGLVEGLRK 3547  
QY 425 RL-----W----- 427  
Db 3548 RLLPANCASLAHGLSLLVAVAVAGVWGVCASFPFGVSVAVLLSSSASFLASFLGWEPLK 3607  
QY 428 ---DVD----- 431  
Db 3608 VLLEALYFSLVAKRLHPDED-DTLVESPAVTPVSARVVRPPHGFALFLAKEARKVKR 3666  
QY 432 ---G-----C-----I---HT---L---T----- 439  
Db 3667 LHGMLRSLVYMLFLVTLTASYGDSCHGHAYRLOSIAKQELHSRAFLAITRSEELWPW 3726  
QY 440 K-----H-Q---E---P---V---Y-----SVA---PS----- 451  
Db 3727 MAHVLLPYVHCNOSPGLPRLRQVRLQALYPPDPGPRVHTCSAAGGFSTDYVGWE 3786  
QY 452 ---PD-----G----- 454  
Db 3787 SPHNGSGTWAYSAPDLLGAWSGWCAVYDSGGYVQELGLSLEESRDLRLFLQLHNWLDNR 3846  
QY 455 ---RY-----L---ASG---S---PD-----K---C 465  
Db 3847 SRAVFLTRYSPAVGLHAAVTLRLBFPAAAGRALAALSVRFPALRRLSAGLSPLTTSVC 3906  
QY 466 ---VH-----I---W-----NT-----Q----- 472  
Db 3907 LLLFAVHFAVAEARTWREGRWVRLGAWARMLLVALTATALVRLAQLGAADROWTRF 3966  
QY 473 ---T-----G---A---LV---H-----S---Y-R----- 481  
Db 3967 VRGRPRFTSFDOQVHSSAARGLAASLLFLLVKAAQHVRFYRQWSVFGKTLRALPEL 4026  
QY 482 -G-T-G---G---IF----- 489  
Db 4027 LGVTLGLVWLVAYAQAILLVSSCVDSLMSVAQALLVLCPTGLSTLCPAESWHLSPLL 4086  
QY 490 C-----WNA-----A-G-----D----- 496  
Db 4087 CVGLNWLRLMGLRGLGAVILRWYHRLRGELYRPAPWPDYEMVELFLRLRLMMGLSKV 4146  
QY 497 ---KV-----G-----A-S-AS-----DG-SV----- 507

Db 4147 KEFRHKVRFEGMEPLFSRSGSKVSPDVPPPSAGSDASHPSTSSSQDGLSVSLGRIGT 4206  
QY 508 -C-----V-----L-----D-----L-RK 514  
Db 4207 RCEPEPSRLQAVPEALLTOPDLNQATVEDVYQLEQQLHSLOGRR 4250  
RESULT 14  
US-09-479-467A-2  
; Sequence 2, Application US/09479467A  
; Patent No. 6723557  
; GENERAL INFORMATION:  
; APPLICANT: Sternberg, Paul W.  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE HOMOLOGS REQUIRED FOR MALE MAT  
; TITLE OF INVENTION: BEHAVIOR IN NEMATODES AND ASSAYS BASED THEREON  
; FILE REFERENCE: 18021-2901B  
; CURRENT APPLICATION NUMBER: US/09/479, 467A  
; CURRENT FILING DATE: 2000-01-06  
; PRIOR APPLICATION NUMBER: 60/115,127  
; PRIOR FILING DATE: 1999-01-06  
; NUMBER OF SEQ ID NOS: 16  
; SOFTWARE: Patent In Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 4303  
; TYPE: PRT  
; ORGANISM: Homo sapiens PKD-1 protein  
US-09-479-467A-2

Query Match 69.4%; Score 2539.5; DB 4; Length 4303;  
Best Local Similarity 9.9%; Pred. No. 1.2e-31;  
Matches 414; Conservative 81; Mismatches 14; Indels 3675; Gaps 356;

QY 1 MSIS-----S---D-----EVN---P----- 10  
Db 72 LDVSHNLLRALDVGLLANLSALAEIDISNNKISTLEEGIFANLNFNLSEINLSGNPFECDC 131  
QY 11 -LVY-----R-----Y----- 15  
Db 132 GLAWLPQWAEQQVVRVQPEATCAGPSLAGOPLIGPLDSCGGEYVACLPDSSGT 191  
QY 16 -----LQ-----E-----SG- 20  
Db 192 VAAVSFAAHEGLQPEACSAFCSTGQGLAALSEQWCLCGAAQSSASFACLSLCSGP 251  
QY 21 -----P-S---H---S---AF-----T---FG--- 29  
Db 252 PAPPAPTCTGPTLLQHVFPASPATLVGPHGLASGLAAPHIAAPLPVTDTRWDFDGS 311  
QY 30 -I-----KSH-----I---S-QSN- 38  
Db 312 AEVDAAGPAASHRYVLPGRHYHTAVIALGAGSALLGTDVQVEAAPALELVCPSSVQSD 371  
QY 39 ---I-N---G-----AL---V---P-----P----- 46  
Db 372 SLDSLSTQNGSGLEAAYSIVALGEEPARAVHPLCPSPDTEIFPGNGHCYRLVVEKAALQ 431  
QY 47 -----AA-----LIS-----I---IQ-K---G---L--- 57  
Db 432 AQEQCOAWAGAAALAMVDSPAVQFLVSRVTRSLDVWIGFSTVQGVVEGPAQGEAFSLES 491  
QY 58 -Q-----IN-----EDG-----YV-E-----AE---V---S- 65  
Db 492 CQNWLPGEHPHATAEHCVRGLGPTGWCNVTDLCSAPHYSVCELOQGGPVQDAENLLVGAPSG 551  
QY 66 ---IN---EDG-----TL---F----- 73  
Db 552 DLQGLPTPLAQDQGLSAPHEPVVMPVPGRLSREAFLTAEFGTQELRPAQLRQVYR 611  
QY 74 -----D-----GR---P---I----- 78  
Db 612 LLSTAGTPENGSEPSRSPDNRTQLAPACMPGGRWCPGANICLPDLASCHPOACANGCTS 671



Db 2828 LVDSNPPFFGYISNTYVSTKVASMAFOTQAGAPIERLASERAITVKVPNNSDWAARGH 2887  
QY 358 - - - - -D- - - - -P- - - - -TG- - - - -N- - - - - - - - - - -LLA- - - - - 365  
Db 2888 RSSANSANSVVQPOASVGAIVTLDSSNPAAGLHLQNLTYLLDGHYLSSEPEPYLAVYLH 2947  
QY 366 S- - - - -CS- - - - - - - - - - -D- - - - - - - - - - - - - - - - 370  
Db 2948 SEPRNEHNCASRRIRPESLQAGADHRPYTFIFSPGSRDPAGSVHNLSSHFRWSALQVS 3007  
QY 371 - - - - -M- - - - -T- - - - -L- 380  
Db 3008 VGLYTSLCQYFSEBDMVWRTGELLPLEETSPOAVCLTRHLTAFGASLFVPPSHRVFVFP 3067  
QY 381 - - - - -D- - - - -N- 388  
Db 3068 EPTADVNYVMLTCAVCLVTYVMAAILHLKLDQDASGRAIPFCGQGRFKYIILVKTG 3127  
QY 389 - 388  
Db 3128 WGRSGTTHAVGIMLYGVDSRSHRHLDCDPAFRHNSLDIFRIATPHSLGSLVKIRVWHD 3187  
QY 389 - - - - -OH- - - - - - - - - - -N- - - - -KE- - - - - - - - - - - 393  
Db 3188 NKGLSPAFLQHVIVRDLQTARGAFVLDNDWLSVETEANGGLVEKEVLAASDAALLRFR 3247  
QY 394 - 399  
Db 3248 LVAELQRFDFKHILSIWDRPSRFRTRIQATCCVLLICFLGANAVYGAVGDSAY 3307  
QY 400 - 402  
Db 3308 STGHVSRSLPSVDTVAVGLVSSVYVYVYLAIFLFRMSKSVAGSPSTPAGQVLDI 3367  
QY 403 - - - - - - - - - - -G- - - - - - - - - - -P- - - - -GT- - - - -NNP 409  
Db 3368 DSCLDSSVLDSSFLTFGLHAEQAFVQMKSDLFLODSKSLVCWPGSGEGLSWPDLSDP 3427  
QY 410 - - - - -NANI- - - - - - - - - - -LAS- - - - -A- - - - -SF- - - - - 420  
Db 3428 SIVGSLRLQARGAAGLGLPEEDGFLSPASPYSPAKSFSASDEDLIOQVLAEGVSPAPT 3487  
QY 421 - - - - -D- - - - -ST- 424  
Db 3488 QDTHMETDLLSSLSSTPGKTELTALQRLGELGPPSPGLNWEQQAARLSRTGLVEGLRK 3547  
QY 425 RL- 427  
Db 3548 RLLPACASLAHGLSLLNVAVAVSGWVGASFPFGVSVAVLLSSASFLASFLGWEPLK 3607  
QY 428 - - - - - - - - - - -DVD- -R 431  
Db 3608 VLLEALYFSLVAKRLHPDED-DTLVSPAVTPVSVRVRPPHGFALFLAKEARKVRK 3666  
QY 432 - - - - -G- - - - - - - - - - -I- - - - -C- - - - - - - - - - -HT- - - - -L- - - - -T- - - - - 439  
Db 3667 LHGMRLSLVYMLFLVLTLLASVGDASCHCHAYRLQSAIKQELHSAFLAIRSEELNWP 3726  
QY 440 K- - - - -H- - - - -Q- - - - -E- - - - -P- - - - -V- - - - - - - - - - -SVA- - - - -FS- - - - - 451  
Db 3727 MAHVLLPYVHGNSPELGPRLRQRLQALYPDPGPRVHTCSAAGGFSTSDYDVGWE 3786  
QY 452 - - - - - - - - - - -PD- 454  
Db 3787 SPHNGSCTWAYSAPDLLGAWSGSCAVYDGGYVQBLGLSLEBSRDLRFLQLHNLWDR 3846  
QY 455 - - - - -RY- - - - - - - - - - -L- - - - -ASG- - - - -S- - - - -PD- - - - - - - - - - -K- - - - -C 465  
Db 3847 SRAVFLFETRYSPAVGHAHVTLRFFPAAGRALAALSVRPPFALRRLSAGLSPLTITSVC 3906  
QY 466 - - - - -VH- - - - - - - - - - -I- - - - -W- - - - -NT- - - - - - - - - - -Q- - - - - 472

Db 3907 LLLFAVFAVAEARTWHREGMRVLRIGAWARWLLVALTAATATVRLAQLGAADROWTRF 3966  
QY 473 - - - - -T- - - - - - - - - - -G- - - - -A- - - - -LV- - - - -H- - - - -S- - - - -Y- - - - -R 481  
Db 3967 VRGRPRFTSFDOVAHVSSAARGLAASLLFLLLVKAAQHVRFVRQMSVFOKTCICALPEL 4026  
QY 482 -G-T-G- - - - -G- - - - - - - - - - -IP- - - - - - - - - - -E- - - - -V 489  
Db 4027 LGVTGLVGLVAYQAAILLVSSCVDSLMSVAQALLVLCPTGLSTLCPAESWHLSPLL 4086  
QY 490 C- - - - -WNA- - - - - - - - - - -A- - - - -G- - - - - - - - - - -D- - - - - 496  
Db 4087 CVGLMALRWGALRLGAVILRWYHALRGELYRPAWEPQDYEMVELFRRLRLMMGLSKV 4146  
QY 497 - - - - -KV- - - - - - - - - - -G- - - - - - - - - - -A- - - - -S- - - - -AS- - - - -DG-SV- - - - - 507  
Db 4147 KEFRHKVRFEGMEPLFSRSGSKVSPDPVPPPSAGSDASHPSTSSQDGLSVSLRLGT 4206  
QY 508 -C- - - - - - - - - - -V- - - - - - - - - - -L- - - - - - - - - - -D- - - - -L- - - - -RK 514  
Db 4207 RCEPEPSRLQAVFEALLTQFDRLNQATEDVYQLEQLHSLQGR 4250

## RESULT 15

US-09-052-469-8  
; Sequence 8, Application US/09052469  
; Patent No. 6380360  
; GENERAL INFORMATION:  
; APPLICANT: Harris et al.  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: One Financial Center  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk, 3.50 inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/052,469  
; FILING DATE: Concurrently herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/422,582  
; FILING DATE: 14-April-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9507766.5  
; FILING DATE: 13-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9411900.5  
; FILING DATE: 14-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB PCT/GB94/02822  
; FILING DATE: 23-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9326470.3  
; FILING DATE: 24-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams, Ph.D., Kathleen M.  
; REGISTRATION NUMBER: 34,380  
; REFERENCE/DOCKET NUMBER: 3265/74165  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 345-9100  
; TELEFAX: (617) 345-9111  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4302 amino acids  
; TYPE: amino acid

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-052-469-8

Query Match      69.4%; Score 2538.6; DB 3; Length 4302;
Best Local Similarity 9.8%; Pred. No. 1.2e-31;
Matches 409; Conservative 87; Mismatches 13; Indels 3674; Gaps 352;

QY 1 MSIS-----S-----D-----FVN-----P-----10
Db 72 LDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEGIFANLNLNLSGNPFECDC 131
QY 11 -LVY-----R-----Y-----15
Db 132 GLAWLPRWABEQVRVQVPEAATCAGPSLAGPGLIGPLLDGCGEEYVACLPDNSSGT 191
QY 16 -----LQ-----E-----SG- 20
Db 192 VAAVSFAAHEGLLOPEACSAFCSTCGGLAALSEQWCLCGAAQPSSASFACLSLCSGP 251
QY 21 -----F-S-----H-----S-----AF-----T-----FG-----29
Db 252 PPPAPTTCRGFTLLQHVFPASPAGATLVGPHGLASGLAAFAAFLPVTATRWDFGDS 311
QY 30 --I-----KSH-----I-----S-QSN- 38
Db 312 AEVDAAGPAASHRYVLPGRYHVTAVALGAGSALLGTDVQVEAAPAALELVCPSSVQSD 371
QY 39 -----I-N-----G-----AL-----V-----P-----P-----46
Db 372 SLDLISIQNRGGSGLEAAYSIVALGEEPARAVHPLCPSPDTEIFPGNGHCYRLVVEKAALQ 431
QY 47 -----AA-----LIS-----I-----IQ-K-----G-----L-----57
Db 432 AQEQCQAWAGAALAMVDSPAVQRLVSRVTRSLDVWIGFSTVGQVEVGPAPQGEAFSLES 491
QY 58 Q-----Q-----YV-E-----AE-----63
Db 492 QCNWLPGEPHPTAEHCVRGLPTGWCMTDLCSPHSVYVCELOPGGPVQDAENLLVGAPSG 551
QY 64 -----V-----S-----I-----66
Db 552 DLQGLPTLAQDGLSAPHEPVEVMVFPGLRLSREAFLLTAEFTQBELRRPAQLRLQVYR 611
QY 67 -----N-----E-----D-----G-T- 71
Db 612 LLTAGTPENGSEPEPSPDNRQTQAPACMPGGRWCPGANICLPDASCHPQACANGCTS 671
QY 72 -----LF-----D-----74
Db 672 GPGLPGAPYALWRELFVSAAGPPAQYSVTLHGQDVLMLPGDLVGLQHDAGPGALLHCSP 731
QY 75 -----G-RP-----77
Db 732 APCHPGQAPYLSANASSWLPPLPAQLEGTWACAPACALRLIAATEQLTVLLGLRPFGLR 791
QY 78 -----I-E-----S-----L-----I-----D-----85
Db 792 MPGRYEVRAEVGVNGVSRHNLSCSFDVVSFVAGLRVYVPAPDRGLRYVPTNGSALVLQVDS 851
QY 86 -----A-----V-MP-----DV 91
Db 852 GANATATARWPGSGVSARFENVCPALVATFVPGCFWEITNDTLFSVVALPWLSEGEHVVDV 911
QY 92 V-----Q-T-----R-----Q-Q-----A-Y-----R-----D 101
Db 912 VVENSASRANLSRLTABEPICGLRATPSPEARVLQGVLYRYSVPVEAGSDMVFRTIND 971
QY 102 KLAQ-----Q-----QAAAA-----AAA-----AS-----Q-Q-----121
Db 972 K--QSLTFQNVFVNIYQSAAVFKLSLTASHNVSNVTYNNVTVERMRMQLQVSTVPA 1029
QY 122 -----G-----SA-K-----N-----GE-----TA-----131
```

Db 2108 PRAHSYLRPGDYRQVQVNASNLVSFFVAQATVTVQVLACREPEVDVVLPLQVLMRRSQRN 2167  
Qy 306 F--H--SA--P--AL--DV-- 314  
Db 2168 YLEAHVDLRDCVYQTEYRWEVYRTASCQPGPARVALPGVDVSRPLVLPRLALPVGH 2227  
Qy 315 -----D-----W-----OS--N-- 319  
Db 2228 YCFVVFVSGDTLQSIQANVTVAPELVLPIEGGSYRVMSTRDLVLDGSESDPNLE 2287  
Qy 320 -----N-----TF----- 322  
Db 2288 DGOPTLSEHWACVASTQREAGCALNFGPRGSSVTIPRERLAAGVEYTFSLTVWKAGR 2347  
Qy 323 -----A-----SC----- 325  
Db 2348 KEATNTQVTLIRGRVPIVLECVSCAKQAQVYEVSRSSVYVLEGRCLNCSSGSKRGWAA 2407  
Qy 326 ---S--T--D-----M-----C--I----- 331  
Db 2408 RTEFNKTLVLDETTSTGTSAGMLVLRGVRDGEYTFTLTVLGRSGBEGCASIRLSP 2467  
Qy 332 -----H-----V-----C--334  
Db 2468 NRPLGGSCRLFLPGAVHALTKVHFECTGWDHDAEDAGAPLVYALLRLRCRQGHCEFCV 2527  
Qy 335 -K-----LG----- 337  
Db 2528 YGSLSSYGAVLPFGRPHPEVGLVAVVQDLGAUVVALNRSIAITLPEPNGSATGLTVW 2587  
Qy 338 -----Q--D-----R-----P----- 341  
Db 2588 LHGLTASVLPGLLRQADPQHVEYSALVTVLNEYERALDVAAEPKHERHQAIRKNIT 2647  
Qy 342 -----I-----K--T-----FQ----- 346  
Db 2648 ETLVSLRVHTVDIOIAAALQAOCMGPSRELVCRCCLKOTLHKLEAMMLILOAETTAGTV 2707  
Qy 347 -----G--H-----T----- 349  
Db 2708 TPTAIGSILNITGDILHLASSDVRAPOPSLGAESPMSVQAYNLTSALMRLMSR 2767  
Qy 350 --NE----- 351  
Db 2768 VLNEEPLTAGBEIQAQKSDPRSLCYGAPGPGCHFSIPEAFSGALANLSDVVQLIF 2827  
Qy 352 -V-----N-----AI--K-----W----- 357  
Db 2828 LVDSNPPPGYISNYTVSTKVSMATQQAQAIPIERLASERAITVKVPNNSDMAARGH 2887  
Qy 358 -----D--P--TG--N-----LLA----- 365  
Db 2888 RSSANSVVPQASVGAVTLDSSNPAAGHLQNLTYLLDGHYLSEEPYLAAYLH 2947  
Qy 366 S-----CS-----D-----D----- 370  
Db 2948 SEPRPNEHNSASRIRPESLOQADHRPYTFIFSPGSRDPAGSYHLNLSHFWSALQVS 3007  
Qy 371 --M--T--L-----KIW--S-----MK-----Q----- 380  
Db 3008 VGLYTSLCQVFSBEDMVMWRTGELLPLEETSPPRQAVCLTRHLTAFGASLFPVPPSHVRFVP 3067  
Qy 381 -----D--N-----C-----V-----H--D--L-----Q----- 388  
Db 3068 EPTADVNYIYMLTCAVCLVTYVMMAAILHKLQDLQDASGRAIPFCQGRFKYIILVKTG 3127  
Qy 389 ----- 388  
Db 3128 WGRSGTTAHVIMLYGVDSRSGHRHLDGDRAPHRNSLDFRIATPHSLGSVWVKIRWHD 3187  
Qy 389 -----QH-----N-----KE----- 393  
Db 3188 NKGLSPAWFLQHVIVRDLQTAQSAFFLVNDWLSVETEANGGLVEKEVLAASDAALLRFR 3247

Qy 394 -----IV-----T--I--K-----W----- 399  
Db 3248 LLVAELQRGFDFDKHIMLSIWDPRPRSRFTIQRATCCVLLI CLFLGANAVWYGAUGDSAY 3307  
Qy 400 -----SPT----- 402  
Db 3308 STGHVSRLSPLSDTVAVGLVSSVVVYVYLAIFLFRMSRSKVAGSPSTPAGQVLDI 3367  
Qy 403 -----G-----P--GT-----NPN- 409  
Db 3368 DSCLDSSVLDSSPLTFSGLHAEAFVGMKSDLFDDSKSLVCWPSGEGTLLSWPDLLSDPS 3427  
Qy 410 --NANL--M-----LAS--A--SF----- 420  
Db 3428 IVGSNLRLQARGAOGHGLGPEEDGFSLASPYSPAKFSASDEDLIQQVLAEGVSSPAPTQ 3487  
Qy 421 -----D-----ST-----V-----R 425  
Db 3488 DTHMETDLLSSLSSTPGEKTTETLALQRLGELGPPPLANWEQQAARLSRTGLVEGLRKR 3547  
Qy 426 L-----W----- 427  
Db 3548 LIPAWCASLAHGLSLLLVAVAVAVGWSGASFPFGVSVAMLLSSSASFLASFLGWEPLKV 3607  
Qy 428 -----DVD-----R- 431  
Db 3608 LLEALYFSLVAKRLHPDED-DTLVESPANTVPSARVPRVPPHGFALFLAKEARKVKRL 3666  
Qy 432 -G-----I-----C-----I--HT--L--T-----K 440  
Db 3667 HGMLRLLVYMLFLLVTLASYGDSCHGHAYRLQSAIKQELHSRAFLAITRSEELWPMW 3726  
Qy 441 -----H--Q--E--P--V-----Y-----SVA--FS----- 451  
Db 3727 AHVLLPYVHGNSSPGLPRLRQVRLEALYPDPGPRVHTCSAAGGFSTSDYDVGWES 3786  
Qy 452 -----PD-----G----- 454  
Db 3787 PHNGSGTWASAPDLLGAMWSGCAVYDSGGYQELGSLSESRDLRFLQLHNLNDNR 3846  
Qy 455 -----RY-----L--ASG-----S--FD-----K--C- 465  
Db 3847 RAVFELTRYSPAVGLHAAVTLRLEFPAAAGRALAALSVRPFALRLSAGLSPLLTSTVCL 3906  
Qy 466 ---VH-----I-----W-----NT-----Q----- 472  
Db 3907 LLFAVHFAVAEARTWHRGRWRVRLGAWARWLLVALTAATLVRLAQLGAADROMTRFV 3966  
Qy 473 -----T-----G--A--A--LV--H-----S--Y--R----- 481  
Db 3967 RGRPRFTSFDQVAHVSSAARGLAASLLFLLLVKAAQHVRVQWVSFGTKLCRALPELL 4026  
Qy 482 G-T-G-----G-----IF-----E-----VC 490  
Db 4027 GVTGLGLVGLVAVQAAILLVSSCVDSLMSVAQALLVLCPTGLTLCPAESWHLSPLLC 4086  
Qy 491 -----WNA-----A--G-----D----- 496  
Db 4087 VGLMALRLWGLALGALGAVILRWRYHALRGELYPAPBPQDYEMVELFLRLRLMWGLSKVK 4146  
Qy 497 ---KV--G-----A--S--AS-----DG--SV----- 507  
Db 4147 EFRHKVRFEGMEPLPSRSSRGSKVSPDVPFPPSAGSDASHPSTSSSQDLGSLVSLGRIGTR 4206  
Qy 508 C-----V-----L-----D-----L--RK 514  
Db 4207 CEPEPSRLQVFEALLTQFDRLNQATEDVYQLEQLHSLQGR 4249

Search completed: January 3, 2005, 15:28:14  
Job time : 100.667 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:26:14 ; Search time 81 Seconds  
(without alignments)

2282.717 Million cell updates/sec

Title: US-09-987-701-2

Perfect score: 3659

Sequence: 1 MSISDEVNPLVRYLQESG.....GDKVCASGDGVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

#### Database :

Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3355.2	91.7	577	17	US-10-723-860-836
2	3265.6	89.2	542	15	US-10-264-049-2892
3	2744	75.0	395	16	US-10-363-829-446
4	2618.9	71.6	5635	15	US-10-451-168-78
5	2614.8	71.5	5636	14	US-10-032-189-128
6	2614.8	71.5	5636	14	US-10-120-801-72
7	2614.8	71.5	5636	14	US-10-023-634-93
8	2614.8	71.5	5636	16	US-10-408-765A-1895
9	2610.5	71.3	4599	15	US-10-464-368-69
10	2610.5	71.3	4599	15	US-10-464-368-70
11	2596.6	71.0	5175	14	US-10-120-801-74
12	2596.6	71.0	5175	14	US-10-369-493-6859
13	2596.6	71.0	5175	14	US-10-369-493-6861

14	2595.3	70.9	5198	14	US-10-120-801-75	Sequence 75, Appl
15	2595.3	70.9	5198	14	US-10-369-493-6858	Sequence 6858, Ap
16	2595.3	70.9	5198	14	US-10-369-493-6860	Sequence 6860, Ap
17	2595.3	70.9	5198	15	US-10-346-863-34	Sequence 34, Appl
18	2589.2	70.8	4183	13	US-10-087-192-672	Sequence 672, Appl
19	2585.9	70.7	4599	17	US-10-479-875-4	Sequence 4, Appl
20	2585.4	70.7	26926	9	US-09-759-508B-2	Sequence 2, Appl
21	2585.4	70.7	26926	17	US-10-723-860-133	Sequence 133, Appl
22	2584.9	70.6	4636	9	US-09-835-996A-33	Sequence 33, Appl
23	2584.8	70.6	4485	16	US-10-741-601-332	Sequence 332, Appl
24	2581.5	70.6	4563	15	US-10-276-774-1723	Sequence 1723, Ap
25	2580.5	70.5	4544	15	US-10-464-368-68	Sequence 68, Appl
26	2580.5	70.5	4544	17	US-10-479-875-6	Sequence 6, Appl
27	2574.9	70.4	4660	15	US-10-464-368-74	Sequence 74, Appl
28	2572.1	70.3	5985	15	US-10-296-734-810	Sequence 810, Appl
29	2570.9	70.3	4545	15	US-10-464-368-67	Sequence 67, Appl
30	2570.9	70.3	4545	15	US-10-464-368-71	Sequence 71, Appl
31	2569.5	70.2	4545	9	US-09-873-403-2	Sequence 2, Appl
32	2569.5	70.2	4545	11	US-09-750-972-2	Sequence 2, Appl
33	2566.6	70.1	4655	16	US-10-741-601-314	Sequence 314, Appl
34	2566.6	70.1	4655	17	US-10-479-875-3	Sequence 3, Appl
35	2560.7	70.0	5376	14	US-10-028-248A-74	Sequence 74, Appl
36	2560.7	70.0	5376	15	US-10-107-782-74	Sequence 74, Appl
37	2560.1	70.0	5701	9	US-09-864-761-37319	Sequence 37319, A
38	2560.1	70.0	5701	14	US-10-029-386-32438	Sequence 32438, A
39	2559.7	70.0	5374	14	US-10-028-248A-75	Sequence 75, Appl
40	2559.7	70.0	5374	15	US-10-107-782-75	Sequence 75, Appl
41	2558.3	69.9	4495	15	US-10-138-588-20	Sequence 20, Appl
42	2554.4	69.8	4753	14	US-10-369-493-5119	Sequence 5119, Ap
43	2553.6	69.8	5405	9	US-09-922-217-1116	Sequence 1116, Ap
44	2553.6	69.8	5405	13	US-10-025-380-1116	Sequence 1116, Ap
45	2553.6	69.8	5405	17	US-10-723-860-1647	Sequence 1647, Ap

#### ALIGNMENTS

#### RESULT 1

US-10-723-860-836  
; Sequence 836, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Ginsburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 836  
; LENGTH: 577  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-723-860-836

Query Match	91.7%	Score	3355.2	DB 17	Length	577
Best Local Similarity	86.2%	Pred. No.	1.1e-48			
Matches	456	Conservative	38	Mismatches	17	Indels 18; Gaps 10
Qy	1	MSISSDEVNPLVRYLQESGFSSAFTGFKHSISQSNINGALVPPAALISITKGLQYV	60			
Db	52	MSITSDEVNPLVRYLQESGFSSAFTGFKHSISQSNINGTLVPPAALISITKGLQYV	111			
Qy	61	EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQTRQAVRDKLAQQQ--AAAAAAA-AA	117			
Db	112	EAEISINEDGTVFDGRPIESLSLIDAVMPDVQTRQAVFKEKLAQQQASAAAAAATAA	171			
Qy	118	ASQ-----QG-S-----AKNGENTANGAHTIANNHTDM-MEVDGDVEIPPNKAVVLR	166			

```
Db 172 ATAATTTAGVSHQNPSPKREATVNGENRAHSV-NNHA-KPMEIDGEVEIPSSKATVLR 229
Qy 167 GHESEVEICAWNPVSDLLASGSDSTARIWNLSNSTSGSTOLVLRHCIREGGQDVPSNK 226
Db 230 GHESEVEICAWNPVSDLLASGSDSTARIWNLSNSTSGSTOLVLRHCIREGGHVDPSNK 289
Qy 227 DVTSLDWNSBGTLLATGSGYDGFARITWKDGNLSTLGOHKGPIFALKWNNKGNFILLSAGV 286
Db 290 DVTSLDWNTGTLATGSGYDGFARITWEDGNLSTLGOHKGPIFALKWNNKGNFILLSAGV 349
Qy 287 DKTIIWDAHTGEAKQFPFHSAPALVDVQSN--TFASCS--TDMCIHVCKLGDORPIKTF 345
Db 350 DKTIIWDAHTGEAKQFPFHSAPALVDVQ--NNTTFASCS--TDMCIHVCKLGDORPVKTF 408
Qy 346 QGHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQDNVHDLQHNKEIYTIKWSPTGPG 405
Db 409 QGHTNEVNAIKWDPGSMLLASCSDDMTLKIWSMKQEVCIHDLQHNKEIYTIKWSPTGPA 468
Qy 406 TNNPNANMLASAFSDSTVRLWDVDRGICHTLTKHOEPVYSVAFSPDGRYLASGSFDKC 465
Db 469 TSNPNSNIMLASAFSDSTVRLWDIERGVCHTTLTKHOEPVYSVAFSPDGRYLASGSFDKC 528
Qy 466 VHIWNTQTGALVHSYRTGGIFEVCWNAAGDKVGASDGSVCVLDLRK 514
Db 529 VHIWNTQSGNLVHSYRTGGIFEVCWNAAGDKVGASDGSVCVLDLRK 577
```

## RESULT 2

```
US-10-264-049-2892
; Sequence 2892, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; PRIOR FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2892
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (45)_
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2892
```

```
Query Match 89.2%; Score 3265.6; DB 15; Length 542;
Best Local Similarity 80.7%; Pred. No. 4e-47;
Matches 442; Conservative 42; Mismatches 20; Indels 44; Gaps 15;
```

```
Qy 5 SDEVNF--LV-----YR-----YLOESGFHSAFTFGIKSHISQSNNG 41
Db 1 TDC--FKESKVDLQSQQKPLRKTPTHTLCGFCGTGLTC-GFHSAGFTFXIESHISQSNNG 57
Qy 42 ALVPPALISITOKGLOYEAEVSIINEDGLTFGRPTIESLIDAVNPVQVQTOQAYRD 101
Db 58 TLVPPALISITOKGLOYEAEVSIINEDGVTFGRPTIESLIDAVNPVQVQTOQAFRE 117
Qy 102 KLAQQQ--AAAAA--AAASQ-----QG-S-----AKNGENTANGBENGHTIANNHTDM 148
Db 118 KLAQQQASAAAAAATAATAATTSAGVSHQNPSPKREATVNGENRAHSV-NNHA-K 175
Qy 149 -MEVDGVEIPPNKAVVLRGHESEVFTICAWNPVSDLLASGSDSTARIWNLSNSTSGST 207
Db 176 PMEIDGEVEIPSSKATVLRGHESEVFTICAWNPVSDLLASGSDSTARIWNLSNSTSGST 235
```

```
Qy 208 QLVLRHCIREGGQDVPSNKDVTSLDWNSBGTLLATGSGYDGFARITWKDGNLSTLGOHKG 267
Db 236 QLVLRHCIREGGHVDPSNKDVTSLDWNTGTLATGSGYDGFARITWEDGNLSTLGOHKG 295
Qy 268 PIPALKWNNKGNFILLSAGVDKTTIIWDAHTGEAKQFPFHSAPALVDVQSN--TFASCS 326
Db 296 PIPALKWNNKGNFILLSAGVDKTTIIWDAHTGEAKQFPFHSAPALVDVQ--NNTTFASCS 354
Qy 327 TDMCIHVCKLGDORPIKTFQGHTNEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQDNVHDL 386
Db 355 TDMCIHVCKLGDORPVKTFQGHTNEVNAIKWDPGSMLLASCSDDMTLKIWSMKQEVCIHDL 414
Qy 387 LQHNKEIYTIKWSPTGPGTNNPNANMLASAFSDSTVRLWDVDRGICHTLTKHOEPVY 446
Db 415 LQHNKEIYTIKWSPTGPA--TNNPNANMLASAFSDSTVRLWDIERGVCHTTLTKHOEPVY 474
Qy 447 SVAFSPDGRYLASGSFDKCVHIWNTQTGALVHSYRTGGIFEVCWNAAGDKVGASDGS 506
Db 475 SVAFSPDGRYLASGSFDKCVHIWNTQSGNLVHSYRTGGIFEVCWNAAGDKVGASDGS 534
Qy 507 VCVLDLRK 514
Db 535 VCVLDLRK 542
```

## RESULT 3

```
US-10-363-829-446
; Sequence 446, Application US/10363829
; Publication No. US20040142331A1
; GENERAL INFORMATION:
; APPLICANT: Jackson, Stuart E.; Lincoln, Stephen E.;
; APPLICANT: Altus, Christina M.; Dufour, Gerard E.;
; APPLICANT: Chalup, Michael S.; Jackson, Jennifer L.;
; APPLICANT: Jones, Anissa L.; Yu, Jimmy Y.;
; APPLICANT: Wright, Rachel J.; Gietzen, Darryl;
; APPLICANT: Liu, Tommy F.; Yap, Pierre E.;
; APPLICANT: Dahl, Christopher R.; Momiya, Monika G.;
; APPLICANT: Bradley, Diana L.; Rohatgi, Sameer D.;
; APPLICANT: Harris, Bernard; Roseberry Lincoln, Ann M.;
; APPLICANT: Gerstein, Jr., Edward H.; Peralta, Careyna H.;
; APPLICANT: David, Marie H.; Panzer, Scott R.;
; APPLICANT: Flores, Vincent Z.; Daffo, Abel;
; APPLICANT: Marwaha, Rakesh; Chen, Alice J.;
; APPLICANT: Chang, Simon C.; Au, Alan P.;
; APPLICANT: Imman, Rebekah R.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PT-1183 USN
; CURRENT APPLICATION NUMBER: US/10/363,829
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: PCT/US01/27628
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/229,751
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,749
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,750
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,747
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,748
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,583
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,517
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,610
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,597
; PRIOR FILING DATE: 2000-09-06
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: PERL Program
```



```

; SEQ ID NO 446
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: LI:021759.1.orf2.2000SEP08
; US-10-363-829-446

Query Match      75.0%; Score 2744; DB 16; Length 395;
Best Local Similarity 99.5%; Pred. No. 5.4e-39;
Matches 389; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSISDEVNVLVRYLQESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db |||||
QY 5 MSISDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISIIQKGLQYV 64
Db |||||
QY 61 EAEVSNEDGTLDFGRPIESLSLIDAVMPDVOTROQAYRDKLAQQAQAAAAAASQ 120
Db |||||
QY 65 EAEVSNEDGTLDFGRPIESLSLIDAVMPDVOTROQAYRDKLAQQAQAAAAAASQ 124
Db |||||
QY 121 QGSAKAGENTANGAHTIANNHTDMEVDGVEIPPNKAVVLRGHESEVEFICAWNPV 180
Db |||||
QY 125 QGSAKAGENTANGAHTIANNHTDMEVDGVEIPPNKAVVLRGHESEVEFICAWNPV 184
Db |||||
QY 181 SLLASGSGDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
Db |||||
QY 185 SLLASGSGDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 244
Db |||||
QY 241 ATGSDGFARITWYKGNLSTLQHGKPIPALKNKKNFILSAGVDKTTIIWDAHTGEA 300
Db |||||
QY 245 ATGSDGFARITWYKGNLSTLQHGKPIPALKNKKNFILSAGVDKTTIIWDAHTGEA 304
Db |||||
QY 301 KQOPPHSAPALVDQSNNTFASCSTDCMHVCKLGQDRPIKTFQGTHTNEVNAIKWDPT 360
Db |||||
QY 305 KQOPPHSAPALVDQSNNTFASCSTDCMHVCKLGQDRPIKTFQGTHTNEVNAIKWDPT 364
Db |||||
QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQHN 391
Db |||||
QY 365 GNLLASCSDDMTLKIWSMKQDNCVHDLQHN 395
Db |||||

RESULT 4
US-10-451-168-78
; Sequence 78, Application US/10451168
; Publication No. US20040091969A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; APPLICANT: GLAXO GROUP LIMITED
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50039
; CURRENT APPLICATION NUMBER: US/10/451,168
; CURRENT FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: PCT/US01/49232
; PRIOR FILING DATE: 2000-12-17
; PRIOR APPLICATION NUMBER: 60/256,710
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/257,048
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/260,482
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/264,922
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/266,797
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/276,988
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/281,535
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/289,622
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 110

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 5635
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-451-168-78

Query Match      71.6%; Score 2618.9; DB 15; Length 5635;
Best Local Similarity 9.6%; Pred. No. 1.3e-29;
Matches 440; Conservative 59; Mismatches 13; Indels 4071; Gaps 372;

QY 1 M-S-----I-----SS---D-----EV----- 8
Db |||
QY 1 MISWEVHTVFLPALLYSLAQDASQSEIRABEIPEGASTLAFVFDVTGSMYDDLQVVI 60
Db |||||
QY 9 -----NF-LV----- 12
Db |||||
QY 61 EGASKILETSLKRKRPLNFALVPHDPFPIGVTITDPKKFOYELRELYVQGGDCPE 120
QY 13 -----Y-----R---Y-----LQ----- 17
Db |||||
QY 121 MSIGAIIKIALEISLPGSFYIVFTDARSKDYRLTHEVLQIIQOKSQVVFVLTGDCDDRTH 180
QY 18 -----E-----SG-F-----S-H-S-----A-----F 26
Db |||||
QY 181 IGYKVEEIASTSSGQVFLDKKQVNEVLKWBVEAQAQSKVHLLSTDLHLEQAVNTWRIPF 240
QY 27 -----T-----FG-----I-----K 31
Db |||||
QY 241 DPSLKEVTVSLSGSPSMIIRNPLGLIKKFGHLHLLNIHNSAKVNVNKEPEAGMWTVK 300
QY 32 ---S-H-----I-----S---Q----- 36
Db |||||
QY 301 TSSSGHRSVRIITGLSIDFRAGFSRKPTLDFKTVSRPVQGIPTVYLLNTSGISTPARID 360
QY 37 -----S-----NI-----N-----G----- 41
Db |||||
QY 361 LLELLISGSSSLKTIPIVKYYPHRKPYGIWNISDFVPPNEAFFLKVGTGYDKDYLQFQVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
Db |||||
QY 421 VFSFSSIVDPAPKVTMPKPTGYVLPQGPVDSLLPFTLSFVRNGVTLGVQDYLKESA 480
QY 48 -----A---L-----I-----S---IQ-----KG----- 56
Db |||||
QY 481 SVNLDIAKVTLSDEGFEICIAVSSAGTGRAQTFDFVSEPPVIOVNNVTVTGERAVLT 540
QY 57 -----L-Q-----Y---V---E----- 61
Db |||||
QY 541 CLIIISAVDYNLTWQRNDRVRLAEPARIRTLANLSLELKSVPENDAGEYHCVSSEGGSS 600
QY 62 -A-----EVS-----NE----- 68
Db |||||
QY 601 AASVFLTVQEPKPVMPKNQSGFTGGSEVSMCSATGYPKPKIAMVTNDMFIVGSHRYRM 660
QY 69 --DGLTF-----D-----G-----R---P-----I--- 78
Db |||||
QY 661 TSDGTLFIKNAAPKADAGIYGCILASNSAGTDKQNSTLYRIEAPKLMVQSELLVALGDITV 720
QY 79 -E---S-----L-S-LI-----DA-----Y----- 87
Db |||||
QY 721 MECKTSGIPPPQVWPKGDLRLRPSTFLIDPLGLLKIQETQDLQDAGDYTCVAINAAGR 780
QY 88 -----M-----P-DV-----VQ-----T----- 94
Db |||||
QY 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPFCYVQGYPEPTIKWRRLDNMPIFSRPFS 840
QY 95 -----R----- 95
Db |||||
QY 841 VSSISQIRTGALFILNLWASDKGTICEAENQFGKIQSETTVTGTGLVAPLIGISPSVAN 900
QY 96 ---QQ-----A-----Y-----R---D---KL--- 103
Db |||||
```

Db 901 VIEGQQLTLPCTLLAGNPPIERRWIKNSAMLLQNPYITVRSQSLHIERVQLQDGGGYTC 960  
Qy 104 ---A--- -Q--Q-- -AA- 109  
Db 961 VASNAVGTNNKTSVVVHVLPTIOHQOQILSTIEGIPVTLPCKASGNPKFSPVIMSKKGL 1020  
Qy 110 ---AAA---A-A--- -A-AA-A- 118  
Db 1021 ISTSAFASAGDSLVSYPGSEGEYVCTAINTAGYAKVKVQLVYVPRVFGDORG 1080  
Qy 119 -SQ- -Q- -GS- 123  
Db 1081 LSQDKPVEISVLAGEEVTLPCEVKSLPPPIITWAKETQLISFPSPRHTFLPSSGSKITET 1140  
Qy 124 ---A--- -K- -N- 126  
Db 1141 RTSDSGYLCVATNIAGNVTAQVKNLVHVPKIQORGPKHLKVQVQGRVDIPCNAGGTPLP 1200  
Qy 127 ---GE--N--T--A--- -N-G- 133  
Db 1201 VITWSKGGSTMLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEILHVQE 1260  
Qy 134 ---E--- -E--N- -GA- -H- 139  
Db 1261 PPTVEDLEPPYNTTFQERVANQRIEFPKPAKGTPTIKWLHNGRELTGREPGISILEDG 1320  
Qy 140 T--IA- -N- -N--HT- -DM- 148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVKDKBEQVTVNSVLLNQLTNL 1380  
Qy 149 ---M--- -M- -E- 150  
Db 1381 FCEVEGTFSPILMYKDNQVQTESSTIQTVNGKILKLPATPEDAGRYSCKAINAGTS 1440  
Qy 151 ---VD--- -G- -DV--E- -I- -P- -PN 160  
Db 1441 QKYFNIDVLVPPTIIGTNFNEVSNLNRDVALEQCVKGTFFPDIIHFKDGKPLFLGDPN 1500  
Qy 161 ---K-A- -K-A- -V- -VL 165  
Db 1501 VELLDRGOVLHLKNARNDKRYOCTVSNAGQAKDKIKTIYNPPSIKGNVTTDISVL 1560  
Qy 166 ---RG--- -RG- -H- 168  
Db 1561 INSLIKLECETRGLPMPAITWYKDGQPIMSSQALYIDKQYLHI PRAQVSDSATYTCV 1620  
Qy 169 ---E-S--- -E- -VFI- -C-A- -W- 177  
Db 1621 ANVAGTAESFHVVDVYVPPMIEGNLATPLNKQVVIASHLSLCKAAGNPSPIILTWLKDV 1680  
Qy 178 ---N--- -N- -P-V--SD- 182  
Db 1681 PVKANDNIRIEAGKKLEIMSAQEI DRGQVVCVATSVAGEKEIKYEVVDVLVPPAIEGGDE 1740  
Qy 183 ---LL--- -LL- -ASGS- -G- -D- -S- 191  
Db 1741 TSYFIVMNNLLEDCHVTGSPPTIMWLKDGQLIDERDGFKILLNGRLKLVIAQAQVSN 1900  
Qy 192 ---TA--- -TA- -R- -I-W 196  
Db 1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSGSLSERVVVVKYKPVALQCIANGIPNPSITW 1860  
Qy 197 ---NL--- -NL- -S 199  
Db 1861 LKDDQPVNTAGNLUKIOSGRVLOIAKTILEDAGRYTCVATNAAGETQOQHILHVHEPPS 1920  
Qy 200 -E- -N- -S- -T- -SGST- 207  
Db 1921 LEDAGKMLNVLVSNPVQLECKAAGNPVPIITWYKDNRLLSGSTMTPLNRCQIIDIES 1980  
Qy 208 -Q- -L- -V- -LR-H- -C- -I- 215  
Db 1981 AQISDAGIKVCVAINSAGATELFYSLQVHVAPISGNSNNMVVVVNNPVRLECEARGIPA 2040

Qy 216 ---R--- -R- -E- -G- -Q- -D- 221  
Db 2041 PSUTWLKDGSPVSSFNGQLVLSGGRILATSAQISDTGRYTCVAVNAAGEKORDIDLRY 2100  
Qy 222 -VPSN- -KO- -VT- -S-L- 231  
Db 2101 YVPFNMGBEQNVSVLISOAVELLCSDAIPPTTLTWLKDGHPLLPKPGLSISENSVLK 2160  
Qy 232 ---D--- -D- -W--N--S- -EG- -TLL--A 241  
Db 2161 IEDAQVQDTRGYTCBATNVAGTEKNVNVNPPNIGSGDELQTLTVIEGNLISLICES 2220  
Qy 242 T--- -GS--Y-D- -G- -F- -A- 249  
Db 2221 SGIPPNLIWKKGSPVLTDMSGRVILSGRQLOISIAEKSDAALYSCVASNAVGTAKK 2280  
Qy 250 ---R--- -R- -I- -WTQD- 256  
Db 2281 EYNLQVIRPTITNSGSHPTTEIIVTRGKSISLECEVQGIPTTPTVWTKDGHPLIKAGVE 2340  
Qy 257 ---NL-AS- -T- -L- -QH- 265  
Db 2341 ILDEGHILQKNIHVSDTRGYCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPENISVVE 2400  
Qy 266 K--- -G--P-I--FALK--W- 274  
Db 2401 KNSVSLTCEASGIPPLSIITWF--KOGWPVLSNSVRILSGRMLRLMOTTMEDAGQVTCV 2458  
Qy 275 ---N--- -N- -K-K- -GN- 279  
Db 2459 VRNAAGEERKIFGLSVLPVPHIVGENTLEDVKYKQSVTLTCEVTGNPVPETIWHKDGQ 2518  
Qy 280 ---P-I- -L- -S- -AGV-- 286  
Db 2519 PLODEAHHIISGGRFLQITNVQVPHGTGRYTCLASSPAGKSRSPSLNPFVSPITAGVGS 2578  
Qy 287 D--K--T- -T- -I- -W- 293  
Db 2579 DGNPEDVTILNSPTSLVCEAYSPPATITWFKDGPLESNRNIRILPGGRTLQILNAQE 2638  
Qy 294 D--- -A- -H- -TG- -EA 300  
Db 2639 DNAGRYSCVATNEAGEMIKHYEVKVIPIPIINKDLWPGLSPEVKIKVNNFTLECEA 2698  
Qy 301 ---K--- -K- -Q- -F 304  
Db 2699 YAIPSASLSWYKDGQPLKSDHVNIAANGHTLQIKEAQISDTGRYTVCVASNIAGEDELDF 2758  
Qy 305 ---P-F- -H- 307  
Db 2759 DVNIQVPPSFQKLWEIGNMLDTRNGEAKDVIINNPFISLYCETNAAPPPTLTWYKDGHPL 2818  
Qy 308 -SA- -P- -A- -LDV- -D- 315  
Db 2819 TSSDKVLILPGRGVLPQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLPVPIIKGANSDL 2878  
Qy 316 ---WQ- -SN- 319  
Db 2879 PEEVTVLVNKSALIECLSSGSPAPRNSWQKQDGPLEDHDKHFLSNRILQILNTOITDI 2938  
Qy 320 ---NT- -PAS--C- -S- 326  
Db 2939 GRVYCAENTAGSACKYFNLVHVPSPVIGPKSENITVVNNPFISLTCEVSGFPPLDLSW 2998  
Qy 327 ---T--- -T- -D- -MCI- 331  
Db 2999 LKNEQPIKLTNTLIVPGGRTLIQIRAKVSDGGEYTCIAINQAGESKKKFSLTVVVPPSI 3058  
Qy 332 --H- -V--C- -K- -L- -GO- 338  
Db 3059 KOHDSLSLVNVNREGTSVLECESNAVPPPVITWYKNGRMITESTHVEILADGQMLHIK 3118

QY 339 -D-----R-----P-IK-----TF----- 345  
Db 3119 KAESDTGVVCRAINVAGRDDKNFLNVVPPSIEGPREVIVETISNPVTLTCDATGI 3178  
QY 346 -----Q-----G-----H-----T----- 349  
Db 3179 PPPTIAWLNKHKRIENSLEVRILSGGSKLIARSQHSNGNYTCIASNMEKAQKYF 3238  
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
Db 3239 LSIQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIOWLKDKPIASGETERIRVS 3298  
QY 361 --G---NLL-A-S-----C-S-----D-----D-----MTL----- 373  
Db 3299 ANGSTLNIYCALTSDTGKYTCVATNPAGBEDRIFNLNVVPTPIRGKDBAEKMLTMDT 3358  
QY 374 ---K-----I-W-----SM-----KQ--D--NCV----- 384  
Db 3359 SINIECRATGTPPPQINWLNKGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVTVCASN 3418  
QY 385 ---H-----D-----L-----Q----- 388  
Db 3419 AGVDNKHYNLOVPAPPNMSMGTEBITVLKGSSTSMACITDGTAPASMAWLDGQPLGL 3478  
QY 389 -----Q-----H-N-----KEI 394  
Db 3479 DAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFLKLVLEPPHINGSEHEI 3538  
QY 395 ---YT-----I---K-W----- 399  
Db 3539 SVIVNNPLELTCIASIPAPKMTWMKGRPLPOTQOVTLGGGEVLRISTAQVEDTGRYT 3598  
QY 400 ---SPTG-----P--CT-----N-----N----- 408  
Db 3599 CLASSPAGDDKEYLVRVHVPNNIAGTDEPRDITVLNRQVLECKSDAVPPVITWLNR 3658  
QY 409 -----P-----N-----ANLM-LAS--A--S--F----- 420  
Db 3659 GERLOATPRVILSGRYLQINNADIGDTANYTCVASNIAGKTREFILTVNVPNNIKG 3718  
QY 421 -----DSTV-----RL-W-D--V-----DR-----G--I----- 433  
Db 3719 POSLVILLNKSTVLECIAEGVPTPRITWRKOGAVLAGNHARYSILENGFLHIQSAHVTD 3778  
QY 434 ---C-----I---H-----TL-----T----- 439  
Db 3779 GRYLCMATNAAGTDRRIDLOVHVPSPAPGPTNMTVINVQTTLACEATGPKPSINWR 3838  
QY 440 K-H-----Q-E-----P-V-----Y-----S 447  
Db 3839 KNHLLNVQNSYRLLSSGSLVILSPSVDATYECTVTNGAGDDKRTVDLTQVQVPS 3898  
QY 448 VA-----F-S-----P-D----- 453  
Db 3899 IADEPTDFLVTHAPAVITCTASGVPPSPSIHWTKNGIRLLPRGDGYRILSSGAIEILATQ 3958  
QY 454 ---GRY-----L-----ASG--S--P 462  
Db 3959 LNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIOQPSELHVLNPNILPCEATGTPSPF 4018  
QY 463 -----D-----KCV----- 466  
Db 4019 ITWQEGINVTSGRNVHVLPSGGLQISRAVREDAGTYMCVQNPAGTALGKILNVQVP 4078  
QY 467 ---H-----I-W----- 469  
Db 4079 PVISPHLKEYVIAVDKPIITLSCADGLPPDITWHKDGRAIVBSIRQVRLSSGLQIAFV 4138  
QY 470 -----N-----T-----Q----- 472  
Db 4139 QPGDAGHYTCMAANVAGSSSTSTKLTVHVPPIRIRSTEGHYTVNENSAQLPCVADGIPTP 4198  
QY 473 ---T-----G-----A-----L-VH- 478

Db 4199 AINWKDNVLLANLLGKYTAEPYGEILENVVLEDSGFTCVANNAAGBDTHTVSLTVHV 4258  
QY 479 -----SV-----R-----G-----T----- 483  
Db 4259 LPTFTTLPDGVSLUNKGBQLRLSCKATGIPLPKLTWTFNNNIIPAHFDSVNGHSELVIERV 4318  
QY 484 ---G-----G-----I-F--E--V-----C----- 490  
Db 4319 SKEDSGTYVCTAENSVGFKAIQGVVVKPPVFKGDPSPSNWIEPLGNAILNCEVKGDP 4378  
QY 491 ---WN-----A-----AGD----- 496  
Db 4379 PTIQNRKGVDIIEISHIRIQLNGSLAIYGTVNEADAGDYTCVATNEAGVVERSMSTLQS 4438  
QY 497 ---K-V-----G-----A-----SAS----- 503  
Db 4439 PPIITLEPVETVINAGGKIILNCQATGEBPQPTITWSRQSHSISWDDRVNVLNSNLSYIAD 4498  
QY 504 ---D-----GSV-----C-V----- 509  
Db 4499 AQKEDTSEFECVARNLMGSLVRVPVIVQVHGFGFSOMSAWRACSVTCGKIQKRSRLCNQ 4558  
QY 510 -L-----DL--R-----K 514  
Db 4559 PLPANGGKPCQSGDLEMRNCQNK 4581

RESULT 5  
US-10-032-189-128  
; Sequence 128, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine B  
; APPLICANT: Shinkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: Macdougall, John R  
; APPLICANT: Smithson, Glenna  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826

; PRIOR FILING DATE: 2001-03-22  
 ; PRIOR APPLICATION NUMBER: 60/279,840  
 ; PRIOR FILING DATE: 2001-03-29  
 ; PRIOR APPLICATION NUMBER: 60/282,981  
 ; PRIOR FILING DATE: 2001-04-11  
 ; PRIOR APPLICATION NUMBER: 60/283,656  
 ; PRIOR FILING DATE: 2001-04-13  
 ; PRIOR APPLICATION NUMBER: 60/309,247  
 ; PRIOR FILING DATE: 2001-07-31  
 ; PRIOR APPLICATION NUMBER: 60/311,754  
 ; PRIOR FILING DATE: 2001-08-17  
 ; PRIOR APPLICATION NUMBER: 60/313,331  
 ; PRIOR FILING DATE: 2001-08-17  
 ; NUMBER OF SEQ ID NOS: 260  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 128  
 ; LENGTH: 5636  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: (3003)  
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: (3041)  
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: (3367)  
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.  
 ; US-10-032-189-128

Query Match 71.5%; Score 2614.8; DB 14; Length 5636;  
 Best Local Similarity 9.5%; Pred. NO. 1.5e-29;  
 Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S-----I-----SS--D-----EV-----8  
 DB 1 MISVEVHTVFLPALLVSSLAQDASPOSEIRAREPPEGASTLAFVFDVTSMDLQVVI 60  
 QY 9 -----NF-LV-----12  
 DB 61 EGAKILETSLKRPKRLFNALVPHFDPEIGPVITITDPKKFOYELRELYVQGGDCPE 120  
 QY 13 -----Y-----R--Y-----LQ-----17  
 DB 121 MSGAIKALEISLPGSFIYFTDARSQYRLTHEVLQLIQKOSQVYVFLTGDCDDRTH 180  
 QY 18 -----E--SG--F-----S-H-S-----A-----F 26  
 DB 181 IGYKVVEIASSTSGQVFLDKKQVNEVLKXVVEAVQASKVHLLSTDHLEQAVNTRWIPF 240  
 QY 27 -----T-----FG-----I-----K 31  
 DB 241 DPSLKEVTVSLGSPSMIEIRNPLGLIKKGFGGLHELLNTHNSAKVNVNKEPEAGMWTVK 300  
 QY 32 ---S--H-----I-----S--Q-----36  
 DB 301 TSSSGRHSVRITGLSTDIFRAGFSRKPTLDFKTVSRPVQGIPTYVLLNTSGISTPARID 360  
 QY 37 -----S-----NI-----N-----G-----41  
 DB 361 LLELLSISGSSLLKTI PVKYYPHRKPYGIWNI SDFVPEAEFFLKVTGYDKDDLQVFEVSS 420  
 QY 42 -----A-----L-----VP-----P-----A 47  
 DB 421 VSFSSIVPDAPKVTMPKTPCYTLQPGQIPCSVDSLIPFTLSFRVNGVTGLGVQYLKESA 480  
 QY 48 -----A--L-----I-----S-----IIQ-----KG-----56  
 DB 481 SVSLDIKVTLSDEGFECIAVSSAGTGRAQTFFDSEPPPVIVQVNNVTVPGERAVLT 540

QY 57 -----L-Q-----Y-----V-----E-----61  
 DB 541 CLIIISAVDYNLTWQRNDRVRLAEPARITLANLSLELXSVKENDAGEVHCWVSSEGSS 600  
 QY 62 -A-----EVS-----NE-----68  
 DB 601 AASVFLTVQEPKPVTPMKNQSPFTGGSEVSIKSGATGYPKPKIAWTNDMFI VGS HRYRM 660  
 QY 69 --DGTLP-----D-----G-----R-----P-----I--78  
 DB 661 TSDGTLFIKNAAPKADAGIYGCCLASAGTDKQNSTILRYIEAPKLMVVQSELLVALGDITV 720  
 QY 79 -E--S-----L--S--LI-----DA--V-----87  
 DB 721 MECKTSGIPPPQVWFKGDELEPSTFLIDPLGLLKIQTODLDAGDYTCVAINAEGR 780  
 QY 88 -----M-----P-DV-----VQ-----T-----94  
 DB 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRLLDNMFI PSRPF 840  
 QY 95 -----R-----95  
 DB 841 VSSISQLRTGALFILNLWASDKGTIYCEAENOFKIQSETTVTVTGLVAPLIGISPVAN 900  
 QY 96 ---QQ-----A-----Y-----R--D--KL--103  
 DB 901 VIEGQQLTLPCTLLAGNPIDERRWIKNSAMLLONPYITVRSDGSLHIERVOLGDGEYTC 960  
 QY 104 ---A-----Q--Q-----AA-----109  
 DB 961 VASNVAGTNNKTTVVVVHVLPTIQHGQILSTIEGIPVTLPCKASGNPKPSVIWKKGEL 1020  
 QY 110 ---AAA--A-A-----A--AA--A-----118  
 DB 1021 ISTSAKFSAGADSLYVVSPEGESEGYCVCTATNTAGAKRKVQLTVYVRPVFGDLRG 1080  
 QY 119 -SQ-----Q-----Q-----GS-----123  
 DB 1081 LSQDKPEISVLAGEBVTLPCEVKSLPPIITWAKETQLISPPSPRHTFLPSGSMKITET 1140  
 QY 124 ---A-----K-----N-----126  
 DB 1141 RTSDSGMLCVATNIAGNVTAQVKNLVHVPKIQGPKHLKVQGVQORVIPCNAOQTPLP 1200  
 QY 127 -----GE--N--T-----A-----N--G-----133  
 DB 1201 VITWSKGGSTMLVDGHHVSNPDGTLSDIQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260  
 QY 134 -----E--N-----GA--H-----139  
 DB 1261 PPTVEDELPYNTTFOERVANQRIEPPCPAKGTPKPTIKLHNGRELITGREPGISILEEG 1320  
 QY 140 T--IA-----N-----N--HT--DM-----148  
 DB 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPPVPIKDKEQSVNVLLNQLTNL 1380  
 QY 149 -----M-----E-----150  
 DB 1381 FCEVEGTSPILIMWYKDNVQVTSSTIQVNVNGKILKLFRATPEDAGRYSCAKINAGTS 1440  
 QY 151 ---VD-----G-----DV--E-----I-----P-----PN 160  
 DB 1441 QKYNIDLVLPPTIIGTNFPKEVSVVLNEDVALEQVKGTPTFPDIHWFKDGKPLFLGDPN 1500  
 QY 161 -----K-A-----V-----VL 165  
 DB 1501 VELLDRGQVHLHLKNARNDKGRYQCTVSNAGAKQAKDIKLTIIYIPPSIKGNVTDTISVL 1560  
 QY 166 -----RG-----H-----168  
 DB 1561 INSILKLECKTRGLPMPALTWTYKDGQPI MSSSQALYIDKGOYLHLPRAQVSDSATYTCVH 1620  
 QY 169 -----E-S-----E-----VFI-----C--A-----W-----177

Db 1621 ANVAGTAERSFHVDVVPVPMIEGNLATPLNKQVIAHSLTECNAAAGNPSILTLKDG 1680  
QY 178 -N- - - - -P-V--SD- 182  
Db 1681 PVKANDNFRIAGGKKLEIMSAQEI DRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740  
QY 183 -LL- - - - -ASGS- - - - -D- - - - -S- 191  
Db 1741 TSYFIVMNNLLBELDCHVTGSPPTIMWLKDGQLIDERDGFILLNGRKLVIQAQVSN 1800  
QY 192 -TA- - - - -R- - - - -I-W 196  
Db 1801 GLYRCMAANTAGDHKEFEVTHVPPTIKSSGLSERVWKYKPVALQCIANGIPNPSITW 1860  
QY 197 -NL- - - - -S- 199  
Db 1861 LKDDQPNTAQNKLKIQSSGRVLOIAKTILEDAGRYTCVATNAAGETQOHIQLHVHEPPS 1920  
QY 200 -E- - - - -N- - - - -T- - - - -SGST- - - - - 207  
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPTIWKXNCLLSGSTSWTFLNRQIIDIES 1980  
QY 208 -Q- - - - -L- - - - -V- - - - -IR-H- - - - -C- - - - -I- - - - - 215  
Db 1981 AQISDAGIYKVAINSAGATELFYSLQVHVAPISGSNNMVAVVNNPVRLECEARGIPA 2040  
QY 216 - - - - -R- - - - -E- - - - -G- - - - -G- - - - -Q- - - - -D- - - - - 221  
Db 2041 PSITWLKDGSPVSSFNGLOVLSGGRILALTSTQISDTGRYTCVAVNAAGEKORDILRV 2100  
QY 222 -VPSN- - - - -KD- - - - -VT- - - - -S-L- - - - - 231  
Db 2101 YVPPNMGEBQNVSVLSQAVELLQSDALPPPTLTLWKDGHPLLPKPGLSISENRSVLK 2160  
QY 232 - - - - -D- - - - -W- - - - -N- - - - -S- - - - -EG- - - - -TLL- - - - - 240  
Db 2161 IEDAQVODTGRYTCATNAGTEKKYNNIWPVNIIGSGDELQTLVIEGNLSLCE 2220  
QY 241 AT- - - - -GS- - - - -Y-D- - - - -G- - - - -F- - - - -A- - - - - 249  
Db 2221 SSGIPPNLIWKKGSPVLTDMSGRVRIISGGRQLQISIAEKSDAALYSCVASNAGTAK 2280  
QY 250 - - - - -R- - - - -I- - - - -WTGDG- - - - - 256  
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKISISECEVOGIPPTVTVMKDGHPHIAKGV 2340  
QY 257 -NL-AS- - - - -T- - - - -L- - - - -GOH- - - - - 265  
Db 2341 EILDEGHILQKNIHVSDDTGRYVCVAVNAGMTDKKYDLSVHAPPSIIGNHRSPEISVV 2400  
QY 266 -K- - - - -G-PI- - - - -FALK- - - - -W- - - - - 274  
Db 2401 EKNSVLTCEASGIPLPSTTF- - - - -KDGWPVLSNSVRILSGGRMLRMOTTMEDAGQYTC 2458  
QY 275 - - - - -N- - - - -K-K- - - - -GN- - - - - 279  
Db 2459 VVRNAGEERKIFGLSVLPVPHVIGENTLEDVKVKEQSVLTCTVGTGNPVPTIWHKDG 2518  
QY 280 - - - - -F-I- - - - -L- - - - -S- - - - -AGV- - - - - 286  
Db 2519 OPLQDEEAHHIISGGRFLQITNVQVPHGTGRYTCCLASSPAGHKRSFSLNVFVSPPTIAGVG 2578  
QY 287 -D--K--T-- - - - - -I-W- - - - - 293  
Db 2579 SDGNPEDVTILNSPTSLVCEAYSYPATITWFKDGTPLSNRNIRILPGGRTLQILNAQ 2638  
QY 294 -D- - - - -A- - - - -H- - - - -TG- - - - -E 299  
Db 2639 EDNAGRYSCVATNAEGEMIKHYKVYPIPIINKDLMGPGSPKEVKIKVNNTLFLECE 2698  
QY 300 A- - - - -K- - - - -Q- - - - - 303

Db 2699 AVAIPASASLWYKDGQPLKSDDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758  
QY 304 F- - - - -P-F- - - - -H- 307  
Db 2759 FDVNIQVPPSFQKLWELIGNMLDTRNGEAKDVIIINNPISLYCECTNAAPPPTLTYKDGHP 2818  
QY 308 -SA- - - - -P- - - - -A- - - - -LDV- - - - -D 315  
Db 2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRVTCAVNEAGEDSLQYDVRVLVPIIEGANS 2878  
QY 316 - - - - -WQ- - - - -SN- - - - - 319  
Db 2879 LPPEVTLVNKGALIBCLSSGSPAPRNSWQKQPLLEDHKKFLNGRILQILNTOITD 2938  
QY 320 -NT- - - - -FAS- - - - -C- - - - -S 326  
Db 2939 IGRYVCVAENTAGSAXKYFNLVNHHVPPSVIGPKSENLTVVNNFISLTCEVSGFPDPDL 2998  
QY 327 - - - - -T- - - - -D- - - - -MCI- - - - - 331  
Db 2999 WLKNXQPIKLNNTLIVPGGRTLQIRAKVSDGGEVTCIAINXAGESKKKFSLTVVYVPS 3058  
QY 332 -H- - - - -V- - - - -C- - - - -K- - - - -L- - - - -GO- - - - - 338  
Db 3059 IKDHSESLSVNVRREGTSVLECESNAVPPPVITWYKGRMITESTHVEILADGQMLHI 3118  
QY 339 -D- - - - -R- - - - -P- - - - -IK- - - - - 343  
Db 3119 KKAESDTCQYVCRAINAGRDDKNFHLNVYVPPSIEGPEREVIVETISNPVTLTCDATG 3178  
QY 344 -T- - - - -F- - - - -Q- - - - -G- - - - -H- - - - -T- - - - - 349  
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGSKLQIARSQSDSGNYTCIASNMEGKAQY 3238  
QY 350 - - - - -N-E-V-NA- - - - -IKM- - - - -D- - - - -P- - - - -T- - - - - 360  
Db 3239 FLISQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTLQIWLKDGKPIASGETERIRV 3298  
QY 361 - - - - -G--NL--A--S- - - - -C--S- - - - -D- - - - -D- - - - -D- - - - -MTL- - - - - 373  
Db 3299 SANGSTLIYGALTSDTGKYTCVATNPAGEEDRIFNLNVYVPTIRGNKDEAKLMTYVD 3358  
QY 374 - - - - -K- - - - -I-W- - - - -SM- - - - -KQ- - - - -D- - - - -NCV- - - - - 384  
Db 3359 TSINIECRXTGTPPPQINMLKNGCLPLPLSSHIRLLAAGQVIRVIRVRAQVSDVAVYTCASN 3418  
QY 385 - - - - -H- - - - -D- - - - -L- - - - -Q- - - - - 388  
Db 3419 RAGVDNKHYNLQVAFPPNMDNSMGTEITVLKSGSTSMACITDGTAPSMAWLRDGOPLG 3478  
QY 389 - - - - -Q- - - - -ANLM-LAS-A- - - - -S- - - - -F- - - - -H-N- - - - -KE 393  
Db 3479 LOAHLTVTHGMVQLLKAEETDSGKYTCIASNEAGEVSKHFIKLVLEPPHNGSEHEE 3538  
QY 394 I- - - - -YT- - - - -I- - - - -K- - - - -W- - - - - 399  
Db 3539 ISVIVNNPLELACIASGIPAPKMTWMMKDGRLPQTDQVQTLGGGEVLRISTAQVEDTGRY 3598  
QY 400 - - - - -SPTG- - - - -P- - - - -GT- - - - -N- - - - - 407  
Db 3599 TCLASSPAGDDKKEYLVRVHVPPNIAGTDEPRDITVLRNRQVLTLECKSDAVPPPVITWLR 3658  
QY 408 N- - - - -P- - - - -N- - - - -ANLM-LAS-A- - - - -S- - - - -F- - - - - 420  
Db 3659 NGERLOATPRVILSGRYLQINNADLGDNTANTCVASNIAAGTKTREFILTVNVPPNIKG 3718  
QY 421 - - - - -DSTV- - - - -RL-W- - - - -D- - - - -V- - - - -DR- - - - -G- - - - -I- - - - - 433  
Db 3719 GPQSVILLNKSTVLECIAGVPTPRITWRKOGAVLGNHARYSILENGFLHQAHTVD 3778  
QY 434 - - - - -C- - - - -I- - - - -H- - - - -TL- - - - -T- - - - - 439  
Db 3779 TGRYLCMATNAAGTDRRIDLQVHVPPSIAPGPTNMTVIVNVQTLTACEATGPKPSINW 3838

```
QY 440 -K-H-----Q-E-----P-V---Y-----446
Db 3839 RKGHLNVQDNQNSYRLSSGLVLIISPSVDDTATYECTVTGAGDDKRTVDLTQVPP 3898
QY 447 SVA-----F-S-----P-D-----453
Db 3999 SIADPTDFLVTKHAPAVICTAGVFPFSLHWTNGIRLLPRGDDGVRILSSGAIELAT 3958
QY 454 ---GRY-----L---ASG--S-461
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIQPPSELHVILNNPILLPCEATGTPSP 4018
QY 462 F-----D-----KCV-----466
Db 4019 FITWQKEGINVTSGRNHAVALPSGGLQISRAVREDAGTYMCVAQNPAQTALGRKILNVQV 4078
QY 467 ---H-----I-W-----469
Db 4079 PPVISPILKEYVIADVXPITLSCDGLPPPDITWHKDGRAIVESIRQVLSGSLQIAF 4138
QY 470 ---N-----T-----Q-----472
Db 4139 VQGDAGHYTCMAANVAGSSSTSTKLVHVPPRIRSTEGHYTVNENSAIILPCVADGIPT 4198
QY 473 ---T-----G-----A-----L-VH 478
Db 4199 PAINWKKDNLNLLGKYTAEPYGELILENVVLEDSGFYTCVANNAAAGBDTHVSLTVH 4258
QY 479 ---S-----Y-----R 481
Db 4259 VLPTFTPLPGDVSLNKGEOURLSCKATGIPLPKLTWTFNNNIIPAHFDSVNGHSELVIER 4318
QY 482 ---CT-----G-----IF-----E-----V--C-----490
Db 4319 VSKEDSTGYCTAENSFGVKATGFFVYVKEPPVFKGDPSPNWIEPLGNNALNCEVKGDP 4378
QY 491 ---WN-----A-----AGD-----496
Db 4379 TPTIQNRKRGVDIEISHIRQLNGSLAIYGTVNEDAGDYTCVATNEAGVVERSMSLTQ 4438
QY 497 ---K-V-----G-----A-----SAS-----503
Db 4439 SPPIITLEPVETVINAGKKIILNCOATGEPOPITWSRQHSISWDVRVNLSSNLYIA 4498
QY 504 ---D-----GSV-----C-V-----509
Db 4499 DAQKEDTSEPECVARNLMGSLVRVPVIVQVHGFSQWSAWRACSVTCGKGIOKRSRLCN 4558
QY 510 --L-----DL--R-----K 514
Db 4559 QPLPANGKPCQGSDDLEMRNCQNK 4582

RESULT 6
US-10-120-801-72
; Sequence 72, Application US/10120801
; Publication No. US20030203843A1
; GENERAL INFORMATION:
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shinkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
```

```
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/285609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 72
; LENGTH: 5636
; TYPE: PRT
; ORGANISM: human
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3003)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3041)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (3367)
; OTHER INFORMATION: Wherein Xaa is any amino acid.
; US-10-120-801-72
```

```
Query Match 71.5%; Score 2614.8; DB 14; Length 5636;
Best Local Similarity 9.5%; Pred. No. 1.5e-29;
Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;
```

```
QY 1 M-S-----I-----SS--D-----EV-----8
Db 1 MISWEVVHTVFLFALLYSLAQDASQSEIRAEFPPEGASTLAFVFDVTCGSMYDDLQVVI 60
QY 9 -----NF-LV-----12
Db 61 EGASKILETSLKRPKPLFNALVPFHDPEIGPVTTITDPKFKQYELRELYVQGGDCPE 120
QY 13 -----Y-----R--Y-----IQ-----17
Db 121 MSIGAIIKIALEISLPGSFYVFTDARSKOYRLTHEVLQIQKQSQVVFVLTGDCDDRTH 180
QY 18 -----E-----SG--F-----S-H-S-----A--F 26
Db 181 IGYVVEEIASTSSGQVFFHLDKKQVNEVLKWEAEVQASKVHLLSTDHLEQAVNTRWIPF 240
QY 27 -----T-----I-----FG-----K 31
Db 241 DPSLKEVTVSLSGSPSMIEIRNPLGKIKKFGELHELLHNHSAKVNVNKEPEAGMWTVK 300
QY 32 ---S--H-----I-----S--Q-----36
Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFKKTVSRPVQGIPTVYLLNTSGISTPARID 360
QY 37 -----S-----N-----G-----41
Db 361 LLELLSISGSSSLKTIPIKYYPHKPYGIWNISDFVPNEAFFLKVTGYDKDDYLFQEVSS 420
QY 42 -----A-----L-----VP-----A 47
Db 421 VSPSSIVDPAPKVTMPEKTPGYVLPQGOIPCSVDLSLLPFTLSFVRNGVTGLGVQYLKESA 480
```

QY 48 ---A---L---I---S---IIQ---KG---56  
Db 481 SVSLDIKAVTLSDGFEYECIAVSSAGTGRAQTFFDVSEPPVQVPPNNVTVTGERAVLT 540  
QY 57 ---L---Q---Y---V---E---61  
Db 541 CLIISAVDYNLTQRDRVRLAEPARITLANLSLELKSVEKFNDAEYHCVSSEGGSS 600  
QY 62 -A---EVS---NE---68  
Db 601 AASVFLTVQEPKVTWPKNQSTFGSEVSMCSATGPKPKIAWTNDMPFVSHRYRM 660  
QY 69 ---DGTLP---G---R---P---I---78  
Db 661 TSDGTLFIKNAAPKDAIGYCLASNSAGTDKQNSTLRYIEAPKLMVVQSELLVALGDIIV 720  
QY 79 -E---S---L---S---LI---DA---V---87  
Db 721 MECKTSGIPPPQVQKFKGDLRLPSTFLIIDPLLLGLLKIOETQDLDAQDYTCVAINAAGR 780  
QY 88 ---M---P---DV---VO---T---94  
Db 781 ATGKITLDVGSPPVFIQEPADVMEIGSNVTLPYVQGYPEPTIKWRRLDNMPFISRPFS 840  
QY 95 ---R---95  
Db 841 VSSISQRLTGALFILNWDKTYICEAENQFGKIQSETTIVTGTIVAPLIGISPSVAN 900  
QY 96 ---QO---A---Y---R---D---KL---103  
Db 901 VIEGOQLTLPCTLLAGNPIPERWRKNSAMLLQNPYITVRSDGSLHIERVOLQDGEYTC 960  
QY 104 ---A---Q---QO---AA---109  
Db 961 VASNAVGTNNKTSVVVHVLTPTIQHQQLLSTIEGIPVTLPCASGNPKPSVTWSKKGEL 1020  
QY 110 ---AAA---A---A---A---AA---A---118  
Db 1021 ISTSSAKFSAGADSLVYVSPGESEGYEYCTATNTAGYAKRKVQLTVYVRPVFGDLRG 1080  
QY 119 -SQ---Q---GS---123  
Db 1081 LSQDKPEISVLAGEVTLPCVEKSLPPPIITWAKETQLISPPSRHTFLPSSMKITET 1140  
QY 124 ---A---K---N---126  
Db 1141 RTSDSGMYLCVATNIAGNVTAVKLVNVHVPKIQRGPKHLKVQVGQRPDIPCAOQTPLP 1200  
QY 127 ---GE---N---T---A---N---G---133  
Db 1201 VITWSKGGSTMLVDGHHVSNPDGTLSDOATPSDAGIYTCVATNIAGTDEITLHVQE 1260  
QY 134 ---E---N---GA---H---139  
Db 1261 PPTVEDLEPPYNTTFQERVANQRIEPPCPAKGTPKPTIKWLHNGRELTGREPGISILEG 1320  
QY 140 T---IA---N---N---HT---DM---148  
Db 1321 TLLVIAVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVFIKDKQVSNVLLNQLTNL 1380  
QY 149 ---M---E---150  
Db 1381 FCEVEGTSPSPIMWYKDNQVVBESSTIQTVNNGKILKLFRAPEADAGYSCKAINAGTS 1440  
QY 151 ---VD---G---DV---E---I---P---PN---160  
Db 1441 OKYFNIDVLVPPTIIGTNFPKEVSVLNRDVALEQVKGTPPDDIHWFKDGKPLFLGDEN 1500  
QY 161 ---K---A---V---VL---165  
Db 1501 VELLDRGOVLHLKNARRNDKGRYQCTVSNAAGQAKDIKLTIVIPPSIKGGNVTTDISVL 1560

QY 166 ---RG---166  
Db 1561 INSILKLECKTRGLPMPAITWYKDGQPIIMSSSQALYIDKQYLHIAQVSDSATYCHV 1620  
QY 169 ---E---S---E---VFI---C---A---W---177  
Db 1621 ANVAGTAEKSFHVDVYVPPMIEGNLATPLNKQVIAHSILTLECAAGNPSPIITLWKDGV 1680  
QY 178 ---N---P---V---SD---182  
Db 1681 PVKANDNFRIEAGGKKLEIMSAQEIQRQVIVCVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740  
QY 183 ---LL---ASGS---G---D---S---191  
Db 1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKDGOLIDERDGFILLNKRKLVIQAQVSNT 1800  
QY 192 ---TA---R---I---W---196  
Db 1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSSGLSERVWVKYKPVALQCIANGIPNSITW 1860  
QY 197 ---NL---S---199  
Db 1861 LKDDQPVTAQGNLKIQSSGRVLQIAKTLLEDAGRYTCVATNAAGETOQHIQLHVHEPPS 1920  
QY 200 -E---N---S---T---SGST---207  
Db 1921 LEDAGKWLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLSGSTSMTFLNRQIIDIES 1980  
QY 208 -Q---L---V---LR---H---C---I---215  
Db 1981 AQTSAGIKYCVAINSAGATFYSLQVHVAPISGSNNMVAVVNNVPRLECEARGIPA 2040  
QY 216 ---R---E---G---G---Q---D---221  
Db 2041 PSLTWLKGSPVSSPENGLOVLGGRIALATSTQISDTGRYTCVAVNAAGEKORDILRV 2100  
QY 222 -VPSN---KD---VT---S---L---231  
Db 2101 YVPPNIMBEEQNVSVLISOAVELLQSDAI PPPTLTWLDGHPHLLKPKGLSISENSVLK 2160  
QY 232 ---D---W---N---S---EG---TLL---240  
Db 2161 IEBAQVQDGRYTCBATNVAGTEKKNVNNVWPPNIGSGDELTQLTVEGNLISLCE 2220  
QY 241 AT---GS---Y---D---G---F---A---249  
Db 2221 SSGIPPPNLIWKKSGSPVLTDMSGRVRLSGGRQLQISIAEKSDAALYSCVASNAVAGTAK 2280  
QY 250 ---R---I---WTGKG---256  
Db 2281 KEYNLOVYIRPTITNSGSHPTLIIIVTRGKSI SLECEVOQIPPTVTWMDGHPHLLKAKGV 2340  
QY 257 ---NL---AS---L---COH---265  
Db 2341 EILDEGHILQLKNIHVSDTGRVYCVAVNAVAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400  
QY 266 -K---G---PT---FALK---W---274  
Db 2401 EKNSVSLTCEASGIPLPSTWF---KDGWPVLSNVRILSGGRMLRLMOTWTBEDAQYTC 2458  
QY 275 ---N---K---K---GN---279  
Db 2459 VVRNAAGEBERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPVETWHKDG 2518  
QY 280 ---F---I---L---S---AGV---286  
Db 2519 OPLQDEAHIIISGGRFLQITNVQVPHGTGRYTCCLASSPAGHKSRSPSLNVFVSPTIAGVG 2578  
QY 287 -D---K---T---I---I---W---293  
Db 2579 SDGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLSNRNIIRILPGRTLQILNAQ 2638  
QY 294 -D---A---H---TG---E---299

Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIIINKGDLWGFLSPKEVKIKVNTLTLECH 2698  
QY 300 A-----K-----Q-----Q----- 303  
Db 2699 AVAIPASLSWYKDGOLKSDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDELD 2758  
QY 304 F-----P-F-----H-----H----- 307  
Db 2759 FDNIOVPPSFQKLWEIGNMLDTRNGEAKVDIINNPIISLYCETNAAPPTLTWYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315  
Db 2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVLVLPPIIEGANS 2878  
QY 316 -----WQ-----SN----- 319  
Db 2879 LPBEVTLVNMKSALIECLSGSPAPRNSWQKQOQLLEDDHKKFLSNGRILQILNTQITD 2938  
QY 320 -----NT-----FAS--C-----S 326  
Db 2939 IGRYVCAENTAGSAKKYFNLNVHVPSPVIGPKSENLTVVNNFISLTCEVSGFPFPPDL 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKNXQPIKLTNTLIVPGGRTLIIRAKVSDGGEYTCIAINXAGESKKKFSLTVVVPPS 3058  
QY 332 H-----V--C-----K-----L--GQ----- 338  
Db 3059 IKDHSLSLVNVREGTSVLECSNAVPPVITWYKNGRMITESTHVEILADGQMLHI 3118  
QY 339 -----D-----R-----P-1K----- 343  
Db 3119 KKAESDGTGYVCRAINAGRDKNFNLNVVPPSIEGPEREVIVETISNPVTLTCDATG 3178  
QY 344 -----F-----Q-----G-----H-----T----- 349  
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGSKLOIARSOHSDSGNYTCIASNMEGKAQKY 3238  
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
Db 3239 FLSIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIOWLKGDKPIASGETERIRV 3298  
QY 361 --G--NLL-A--S-----C--S-----D-----D-----MTL-- 373  
Db 3299 SANGSTLIYGTSTGKTYCVATNPAGEEDRIENLNVVTPTRGNKDEAKLMTYVD 3358  
QY 374 -----K-----I-W-----SM-----KO--D--NCV-- 384  
Db 3359 TSINIECRXTGTPPPQINLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418  
QY 385 -----H-----D-----L-----Q----- 388  
Db 3419 RAGVDNKHYNLQVAPPNMDNSMGTEBITVLKGSSTSMACITDGTAPSNMRLDGOPLG 3478  
QY 389 -----Q-----H-N-----KE 393  
Db 3479 LDAHLTVSTHGMVLQLLKAETSDSGKTYCTIASNEAGEVSKHFIKVLPPHINGSBEHEE 3538  
QY 394 I-----YT-----I--K--W----- 399  
Db 3539 ISVIVNNPLETCTIASGIPAKMTWMDGRPLQTDQVQVLGGGEVLRISTAQVEDTGRY 3598  
QY 400 -----SPTG-----P--GT-----N----- 407  
Db 3599 TCLASSPAGDDKEYLVRVHVHPNIACTDPRDITVLRNRQVLTLECKSDAVPPPVITWLR 3658  
QY 408 N-----P-----N-----ANLM-LAS--A--S--F----- 420  
Db 3659 NGERLOATPRVILSGRYLOINNADLGTANTVCVASNIAGKTTREFTLVNVPPIKNG 3718  
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433

Db 3719 GPQSLVILINKSTVLECIAEAGVPTPRIITWRKDGAVLGNHARYSILENGFLHIQSAHVTD 3778  
QY 434 -----C-----I-----H-----TL-----T----- 439  
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHVPPIAPGPTNMVTIVNVQTTLACEATGIPKPSINW 3838  
QY 440 -K--H-----Q--E-----P--V-----Y----- 446  
Db 3839 RONGHLLNVQONQSVRLSSGSLVILSPSVDTTATYECTVTNGAGDDKRTVDLTQVQPP 3898  
QY 447 SVA-----F--S-----P--D----- 453  
Db 3899 SIADSEPTDLVTKHAPAVITCTASGVPPFPIHWTNKGIRLLPRGQGYRILSSGAIFILAT 3958  
QY 454 -----GRY-----L-----ASG--S-- 461  
Db 3959 QLNHAGRYTCVARNAGSAHRHTLVHVEPPVIOQPSLHVILNPPILLPCEATPTSP 4018  
QY 462 F-----D-----KCV----- 466  
Db 4019 FITWQKEGINVNTSGRHHAVLPSSGGLQISRAVREDAGTYMCVAQNPAGTALGKILNVQV 4078  
QY 467 -----H-----I--W----- 469  
Db 4079 PPVISPFLKEYVIAVDKPIITLSCADGLPPDITWKKDGRAIVESIRQVLSGSLQIAF 4138  
QY 470 -----N-----T-----Q----- 472  
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTVHVPPIRSTEGHYTVNENSOAILPCVADGIFT 4198  
QY 473 -----T-----G-----A-----L-VH 478  
Db 4199 PAIWKKNVLLANLLGKYTAEPYBELILENVLEDSGFYTCVANNAAGEDTHVSLTVH 4258  
QY 479 -----S-----Y-----R 481  
Db 4259 VLPTFTPELPGDVLNKGEOQLRSLCKATGIPKLTWTFNNIIPAHFDSVNGHSELVIER 4318  
QY 482 -----GT-----G-----G-----IF-----E--V--C----- 490  
Db 4319 VSKBSGTYVTAENSVGPFVKATGFVYVKEPPVFKGDYPSNWIPLGGMALNCEVKGDP 4378  
QY 491 -----WN-----A-----AGD----- 496  
Db 4379 TPTIQNRKGVDEIESHRIQLNGSLAIVGTWNEDAGDYTCVATNEAGVVERMSLTLQ 4438  
QY 497 -----K-V--G-----A-----SAS----- 503  
Db 4439 SPPIITLEPVTVINAGGKIILNCQATGEPQPTITWSROGHSISWDDRVNLSNLSYIA 4498  
QY 504 -----D-----GSV-----C-V----- 509  
Db 4499 DQAKEDTSFECVARNLMGSLVVRVPIVQVHGGFSQSAWRACSVTCGKGIQKRSCLN 4558  
QY 510 --L-----DL--R--K 514  
Db 4559 QPLPANGKPKCGSDLEMCNQNK 4582

## RESULT 7

US-10-023-634-93

: Sequence 93, Application US/10023634

: Publication No. US20030236389A1

: GENERAL INFORMATION:

: APPLICANT: Shimkets, Richard A

: APPLICANT: Colman, Steven D

: APPLICANT: Spytek, Kimberly A

: APPLICANT: Ballinger, Robert A

: APPLICANT: Guo, Xiaojia

: APPLICANT: Tchernev, Velizar T

: APPLICANT: Shenoy, Suresh G

: APPLICANT: Li, Li

: APPLICANT: Ellerman, Karen



APPLICANT: Zerhusen, Bryan D  
APPLICANT: Patturajan, Meera  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc  
APPLICANT: Gusev, Vladimir Y  
APPLICANT: Burgess, Catherine E  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Gangolli, Esha A  
APPLICANT: Malyankar, Uriel M  
APPLICANT: Gunther, Erik  
APPLICANT: Smithson, Glenda  
APPLICANT: Millet, Isabelle  
APPLICANT: Gerlach, Valerie  
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
FILE REFERENCE: 21402-221  
CURRENT APPLICATION NUMBER: US/10/023,634  
CURRENT FILING DATE: 2002-06-28  
PRIOR APPLICATION NUMBER: 60/256,025  
PRIOR FILING DATE: 2000-12-15  
PRIOR APPLICATION NUMBER: 60/265,163  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 60/272,929  
PRIOR FILING DATE: 2001-03-02  
PRIOR APPLICATION NUMBER: 60/274,864  
PRIOR FILING DATE: 2001-03-09  
PRIOR APPLICATION NUMBER: 60/276,688  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/277,880  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/286,409  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/309,246  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/315,600  
PRIOR FILING DATE: 2001-08-29  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 93  
LENGTH: 5636  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (3003)  
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the  
OTHER INFORMATION: specification.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (3041)  
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the  
OTHER INFORMATION: specification.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (3367)  
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the  
OTHER INFORMATION: specification.  
US-10-023-634-93

Query Match 71.5%; Score 2614.8; DB 14; Length 5636;  
Best Local Similarity 9.5%; Pred. No. 1.5e-29;  
Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S-----I-----SS---D-----EV-----8  
DB 1 MISWEVHTVFLPALLYSLAQDASPQSEIRABEPPEGASTLAFVFDVTGSMYDDLQVI 60  
QY 9 -----NF-LV-----12  
DB 61 EGASKILETSLRPKPLNFALPHDPHDPGIVTITDPKQFQELRELYVGGGDCPE 120  
QY 13 -----Y-----R---Y-----LQ-----17

DB 121 MSIGAIAKIALEISLPGSFYVFTDARS KDYRLTHEVLQLIQKQSQVFLVLTGDCDDRTH 180  
QY 18 -----E-----SG--F-----S-H-S-----A-----F 26  
DB 181 IGYKVEEIASTSSGQVFLDKKQVNEVLKQVBEAQAASKVHLLSTDHLEQAVNTWRIPF 240  
QY 27 -----T-----FG-----I-----K 31  
DB 241 DPSLKEVTVSLSGSPSMIBIRNPLGKLKKGFGLHELLNIHNSAKVNVNKEPEAGMTVK 300  
QY 32 ---S-H-----I-----S-Q-----36  
DB 301 TSSGRHSVRITGLSTIDFRAGSRKPTLDFKTKTSRVPVQGIPTYYLLNTSGISTPARID 360  
QY 37 -----S-----NI-----N-----G-----41  
DB 361 LLELLSISGSSSLKTIPIVKYYPHRKPYGIWNIISDFVPPNEAPFLKVTGYDKDYLFORVSS 420  
QY 42 -----A-----L-----VP-----P-----A 47  
DB 421 VSFSSIVDPAPKVTMEKTPGVYLPQGPICSVDSLPFTLSFVRNGVTGLGVQYLKESA 480  
QY 48 -----A--L-----I-----S--IIQ-----KG-----56  
DB 481 SVSLDIKVTLSDEGFYECIAVSSAGTGRAOTFFDVSEPPVVIQVPPNVTVTGERAVLT 540  
QY 57 -----L-Q-----Y--V-E-----61  
DB 541 CLIIISAVDYNLTWQRNDRVRLAEPARTLANLSLELSKVKENDAGEYHCVMSSEGSS 600  
QY 62 -A-----EVS-----NE-----68  
DB 601 AASVFLTVQEPKVTVMKQSFSGSEVSIKMSATGYPKPKIAWTNDMFI VGS HRYM 660  
QY 69 --DGLTF-----D-----G-----R--P-----I--78  
DB 661 TSDGTLFIKNAAPKAGIYGCLASNSAGTDKONSTLRYIEAPKLMVQSELLVALGDITV 720  
QY 79 -E--S-----L--S-LI-----DA--V-----87  
DB 721 MECKTSGIPPPQVKWPKGDLRLPSTFLIIDPLGLLKIQETQDL DAGDYTCVAINAAGR 780  
QY 88 -----M-----P-DV-----VQ-----T-----94  
DB 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRLLDNMFI SRPFS 840  
QY 95 -----R-----95  
DB 841 VSSISQLRTGALFILNLWASDKGTVICEAENOFKIQSETTVTVTGLVAPLIGISPSVAN 900  
QY 96 ---Q-----A-----Y-----R--D--KL--103  
DB 901 VIEGQQLTLPCTLLAGNPPIERRWIKNSAMLLQNPIYTVRS DGSGLHIERVQLDGGEYTC 960  
QY 104 ---A-----Q-QQ-----AA-----109  
DB 961 VASNVAGTNKTTSVVVHVLPTIQHCQQLSTIEGIPVTLPCCKASGNPKPSVIWKKGEL 1020  
QY 110 ---AAA--A-A-----A-AA--A-----118  
DB 1021 ISTSAKFSAGADGSLVYVSPGESESGEYVCTATNTAGAKRKVQLTVYVPRVFGDLRG 1080  
QY 119 -SQ-----Q-----GS-----123  
DB 1081 LSQDKPVEISVLAGEBVTLPCEVKS LPPPIITWAKETQLISPFSPRHTFLPSGSMKITET 1140  
QY 124 ---A-----K-----N-----126  
DB 1141 RTSDSGMVLCVATNAGNVTQAVKLVNHYVPKIQGRPKHLKVQGVQGVQVPCNAOQTPLP 1200  
QY 127 -----GE--N--T-----A-----N-G-----133  
DB 1201 VITWSKGGSTMLVDGEHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDEBTILHVQE 1260

QY 134 -----E-----E-----N-----GA-----H----- 139  
Db 1261 PPTVEDLEPPYNTTFOERAVANQRIEFPCPAKGTPKPTIKWLHNGRELTGREGPISILEEG 1320  
QY 140 T---IA-----N-----N-----HT-----DM----- 148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPPVVKDEQVSNVSVLLNQLTNL 1380  
QY 149 -----M-----E----- 150  
Db 1381 FCEVEGTPSPIMYKDNQVNTESSTIQTNNNGKILKLFATPDAGKYSCKAINIAGTS 1440  
QY 151 -----VD-----G-----DV-----E-----I-----P-----PN 160  
Db 1441 QYFNIDVLVPTTIIGTNFPKESVVLNRDVALSCQVKGTPFPDIHWFKDGKPLFLGDPN 1500  
QY 161 -----K-----A-----V-----VL 165  
Db 1501 VELLDRGOVLHLKNARNDKGRYQCTVSNAGKQAKDKILTIYIPPSIKGNVTTDISVL 1560  
QY 166 -----RG-----H----- 168  
Db 1561 INSLIKLECKTRGLPMPAITWYKDGQPTMSSQALYIDKGQYLHPRAQVSDSATYTCV 1620  
QY 169 -----E-----S-----VFI-----C-----A-----W----- 177  
Db 1621 ANVAGTAKSPHVDVYVPPMIEGNLATPLNKQVVIASHLTLECNAGNPPSILTWLKDGV 1680  
QY 178 -----N-----P-----V-----SD----- 182  
Db 1681 PVKANDNFRIEAGCKLEIMSAQELDRGOYICVATSVAGEKEIKYEVVDVLVPAIEGDE 1740  
QY 183 -----LL-----ASGS-----G-----D-----S----- 191  
Db 1741 TSYFIVMNNLELDCHVTGSPPTIMWLKDGQLIDERDGFKILLNGKLVIAQAQVSN 1800  
QY 192 -----TA-----R-----I-----W 196  
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPTIKSGLSERVVVVKYKVALQCIANGIPNPSITW 1860  
QY 197 -----NL-----S-----S 199  
Db 1861 LKDPQPVNTAQNLKIQSSGRVLQIAKTLLLEDAGRYTCVATNAAGETQQHIQLHVHEPPS 1920  
QY 200 -E-----N-----S-----T-----SGST----- 207  
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLSGTSMTFLNRGQIIDIES 1980  
QY 208 -Q-----L-----V-----LR-----H-----C-----I----- 215  
Db 1981 AQISDAGIYKCVAINSAGATELFYSQVHVAPSISGNSNMVAVVNNPVRLCEARGIPA 2040  
QY 216 -----R-----E-----G-----Q-----D----- 221  
Db 2041 PSLTWLKDGSVPSSFSNGLQVLSGRILALTSTQISDTGRYTCVAVNAAGEKQDIDLVR 2100  
QY 222 -VPSN-----KD-----VT-----S-----L----- 231  
Db 2101 YVPPNIMGEEQNVSVLISQAVELLQSDAIPPTTLTWLKDGHPLKPKGLSISENRSVLK 2160  
QY 232 -----D-----W-----N-----S-----EG-----TL----- 240  
Db 2161 IEDAQVODTGRYTCVATNAGTKKYNVNIWPPNIGGSDTLTQLTVEGNLISLCE 2220  
QY 241 AT-----GS-----Y-----D-----G-----F-----A----- 249  
Db 2221 SSGIPPNLWKKKGSPLVTDMSGRVRLSGGRQLQISIAEKSDAALYSVAVNAGTAK 2280  
QY 250 -----R-----I-----WTKDG----- 256  
Db 2281 KEYNLQYIRPTITNSGHPTEIIVTRKGSISLECEVOGIPPTVTVMKDGHPLIKAGV 2340

QY 257 -----NL-----AS-----T-----L-----GOH----- 265  
Db 2341 EILDEGHILQKNHISDTRGYVCVAVNAGMTKKYDLSVHAPPSIIGNHRSPENISVV 2400  
QY 266 -K-----G-----PI-----FALK-----W----- 274  
Db 2401 EKNSVSLTCEASGIPLPSTIWF--KDGWPVLSLSNVRILLSGGRMLRLMQTTMEDAGQYTC 2458  
QY 275 -----N-----K-----K-----GN----- 279  
Db 2459 VVRNAAEERKIFGLSVLPPHIVGENTLEDVVKQKQSVTLTCEVTGNPVEITWHKOG 2518  
QY 280 -----F-----I-----L-----S-----AGV----- 286  
Db 2519 QPLODEAHHIISSGRFLQITNVQPHTRGYTCCLASSPAGKSRKPSLNVFVSPTIAGVG 2578  
QY 287 -D-----K-----T-----I-----I-----W----- 293  
Db 2579 SDGNPEDVTILNSPTSLVCEAVSYPPATITWFKDGTPLSNRNIRILPGRPTLQILNAQ 2638  
QY 294 -D-----A-----H-----TG-----E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIIINKGDLWGPGLSPKEVKIKVNNTLLECE 2698  
QY 300 A-----K-----Q-----O----- 303  
Db 2699 AYAIPSASLSWYKDGOPKSDDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758  
QY 304 F-----P-----F-----H----- 307  
Db 2759 FQVNIQVPPSFQKWEIGNMLDTRNGEAKDVIINNPIISLYCETNAAPPTTLTWYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315  
Db 2819 LTSSDKVILIPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878  
QY 316 -----WQ-----SN----- 319  
Db 2879 LPBEVTVLVNKSALIECLSSGSPAPRNSWQKQOPLEDHDKHFLNGRILQILNTQITD 2938  
QY 320 -----NT-----FAS-----C-----S 326  
Db 2939 IGRYVCVAENTAGSAKKYFNLNVHVPPSVIGPKSENLTVVNNFISLTCEVSGFPDPDLS 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKNXQPIKLTNTLIVPGGRTLQIIIRAKVSDGGEYTCIANKAGESKKKFSITVYVPPS 3058  
QY 332 --H-----V-----C-----K-----L-----GO----- 338  
Db 3059 IKDHDSESLSVNVREGTSVLECESNAVPPPVITWYKGRMITESTHVEILLADGQMLHI 3118  
QY 339 -----D-----R-----P-----IK----- 343  
Db 3119 KKAESVDTGOYVCRAINVAGRDDKNFHLNVVPPSVIEGPEREVIVETISNPVTLTCDATG 3178  
QY 344 -----T-----P-----Q-----G-----H-----T----- 349  
Db 3179 IPPPTTIAWLKNYKRIENSLSLEVRILSGGSKLQIARSQHSDSGNYTCIASNMEGKAQY 3238  
QY 350 -----N-----E-----V-----NA-----IKW-----D-----P----- 360  
Db 3239 FLISQVPPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLIQLWKDGKPIASGETERIRV 3298  
QY 361 -----G-----NLL-----A-----S-----C-----S-----D-----MTL----- 373  
Db 3299 SANGSTLNIYALTSDTGRYTCVATNPAGEBRIFNLVNVVTPTRIGNKDEAKLMTYVD 3358  
QY 374 -----K-----I-----W-----SM-----KQ-----D-----NCV----- 384  
Db 3359 TSINIECRXTGTPPPPOINLKNGLPLPLSSHRLLLAAGQVIRIVRAQVSDVAVYTCVSN 3418  
QY 385 -----H-----D-----L-----Q----- 388

```
Db 3419 RAGVDNKHYNLQVAFPPNMDNSMGTEBITVLKSGTSMACITDGTAPSMAWLRDQPLG 3478
QY 389 - - - - - O - - - - - H-N - - - - - KE 393
Db 3479 LDAHLLVTHSGMWLQLLKARTDSGKYCTCIASNEAGEVSKHFLKVLPPHNGSEHEE 3538
QY 394 I - - - - - Y - - - - - I - - - - - K - - - - - W - - - - - 399
Db 3539 ISVIVNNPLELTCIASGIPAPKWTMMDKGRPLQTDQVQTLGGGEVLRISTAQVEDTGRY 3598
QY 400 - - - - - SPTG - - - - - P - - - - - GT - - - - - N - - - - - 407
Db 3599 TCLASSPAGDDKEYLVRHVHPNIAGTDRPRDITVLENRQVTLCKSDAVPPPVITWLR 3658
QY 408 N - - - - - P - - - - - N - - - - - ANLM-LAS - - - - - A - - - - - S - - - - - F - - - - - 420
Db 3659 NGERLOATPRVILSGRYLOINNALDGTANTYTCVASNIAGKTTREFFILTVNVPNIK 3718
QY 421 - - - - - DSTV - - - - - RL-W - - - - - D - - - - - V - - - - - DR - - - - - G - - - - - I - - - - - 433
Db 3719 GPQSLVILLKSVLECIASGVPTPRITWRKDGAVLAGNHARYSILENGFLHIQSAHVTD 3778
QY 434 - - - - - C - - - - - I - - - - - H - - - - - I - - - - - TL - - - - - T - - - - - 439
Db 3779 TGRYLCMATNAAGTDRRRIDLVHVPPSIAPGPTNMTVIVNQVTLTACEATGIPKPSINW 3838
QY 440 - - - - - K - - - - - H - - - - - Q - - - - - E - - - - - P - - - - - V - - - - - Y - - - - - 446
Db 3839 RKNGLHLLNDQNSYLLSSGLVLIISPSVDDTATYECTVTNGAGDKRTVLDLTQVQVP 3898
QY 447 SVA - - - - - - - - - - F - - - - - S - - - - - P - - - - - D - - - - - 453
Db 3899 SIADPTDFLVTKHAPAVICTASGVFPFSIHWTXKNGIRLLPRGDGYRILSSGAIBLAT 3958
QY 454 - - - - - GRV - - - - - - - - - - L - - - - - ASG - - - - - S - - - - - 461
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLVHVEPPVIOPOPSSELVHLNPNILLPCEATGTPSP 4018
QY 462 F - - - - - - - - - - D - - - - - KCV - - - - - 466
Db 4019 FITWQEGINVTSGRNHAVLPSGGLQISRAVEDAGTYMCVAQNPAGTALGKIKLVQV 4078
QY 467 - - - - - H - - - - - - - - - - I - - - - - W - - - - - 469
Db 4079 PPVISHLKEYVIAVDKPTLSCAADGLPPDITWHKDGRAIVESIRQVLSGSLQIAF 4138
QY 470 - - - - - N - - - - - - - - - - T - - - - - Q - - - - - 472
Db 4139 VQPDAGHYTCMAANVAGSSSTKLTVHVPPRIRSTEGHYTVNENSQAILPCVADGPT 4198
QY 473 - - - - - T - - - - - - - - - - G - - - - - A - - - - - L-VH 478
Db 4199 PAINWKNVLLANLLGKYTAEPYGBELILENVVLEDSGFYTCVANNAAGEDTHVSLTVH 4258
QY 479 - - - - - S - - - - - - - - - - Y - - - - - R - - - - - 481
Db 4259 VLPTFTELPDGSVNLNKGELRLSKATGIPLPKLTWTFNNIIPAHFDSVNGHSELVIER 4318
QY 482 - - - - - GT - - - - - G - - - - - IF - - - - - E - - - - - V - - - - - C - - - - - 490
Db 4319 VSKEDSGTYVCTAENSVGFKATGFVVVKEPPVFKGDYPSNWIPLGGNAILNCEVKGDP 4378
QY 491 - - - - - WN - - - - - A - - - - - AGD - - - - - 496
Db 4379 TPTIOWNRKGVDIIEISHRIQLNGSLAIYGTVNEDAGDYTCVATNEAGVVERMSLTQ 4438
QY 497 - - - - - K-V - - - - - G - - - - - A - - - - - SAS - - - - - 503
Db 4439 SPPIITLEPVETVINAGKILNCOATGEPOPTITWSRQGHISWDNRVNLNNSLYTA 4498
QY 504 - - - - - D - - - - - GSV - - - - - C-V - - - - - 509
```

```
Db 4499 DAQKEDTSBEFCVARNLMGSLVLRVPVIVQVHGFGFSQWSAMRACSVTCGKGIQKRSLCN 4558
QY 510 - - - - - L - - - - - DL - - - - - R - - - - - K 514
Db 4559 QPLPANGGKPCQSGDLEMCNCQNK 4582

RESULT 8
US-10-408-765A-1895
; Sequence 1895, Application US/10408765A
; Publication No. US20040101874A1
; GENERAL INFORMATION:
; APPLICANT: Ghosh, Soumitra S.
; APPLICANT: Fahy, Eoin D.
; APPLICANT: Zhang, Bing
; APPLICANT: Gibson, Bradford W.
; APPLICANT: Taylor, Steven W.
; APPLICANT: Glenn, Gary M.
; APPLICANT: Warnock, Dale E.
; TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
; FILE OF INVENTION: IDENTIFIED IN THE MITOCHONDRIAL PROTEOME
; FILE REFERENCE: 560088.465
; CURRENT APPLICATION NUMBER: US/10/408,765A
; NUMBER OF SEQ ID NOS: 3077
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1895
; LENGTH: 5636
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 3003, 3041, 3367
; OTHER INFORMATION: Xaa = Any Amino Acid
US-10-408-765A-1895
```

```
Query Match 71.5%; Score 2614.8; DB 16; Length 5636;
Best Local Similarity 9.5%; Pred. No. 1.5e-29;
Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S - - - - - I - - - - - SS - - - - - D - - - - - EV - - - - - 8
Db 1 MISWEVHTVFLFALLYSLAQDASPSQSEIRABEFPPEGASTLAFVFDVTGSMYDDLQVVI 60
QY 9 - - - - - - - - - - NF-LV - - - - - 12
Db 61 EGASKILETSKRKRPLNFALVPHDPDEIGPVTTTDPKPFQYELRELYVGGGDCPE 120
QY 13 - - - - - Y - - - - - R - - - - - Y - - - - - LQ - - - - - 17
Db 121 MSIGAIIKIALEISLPGSFYVFTDARSKDYRLTHEVLQLIQKQSQVVFVLTGDCDDRTH 180
QY 18 - - - - - E - - - - - SG - - - - - F - - - - - S - - - - - H - - - - - S - - - - - A - - - - - P 26
Db 181 IGKVVYEETASTSGQVPHLDKKQNVNVLKWEAAVQASKVHLLSTDHLEQAVNTWRIFP 240
QY 27 - - - - - T - - - - - - - - - - FG - - - - - I - - - - - 31
Db 241 DPSLKEVTVSLGSPSMIEIRNPLGKLIKGFGLHLLHNSAKVNVNKEPEAGMWTVK 300
QY 32 - - - - - S - - - - - H - - - - - I - - - - - S - - - - - Q - - - - - 36
Db 301 TSSSGRHSVRITGLSTIDFRAGFSRKPTLDFFKTVSRPVQGIPTYVLLNTSGISTPARID 360
QY 37 - - - - - S - - - - - - - - - - NI - - - - - N - - - - - G - - - - - 41
Db 361 LLELLISGSSSLKTIPTKYYPHRKPYGIWNISDFVPPNEAFFLKVTGYDKDYLFQVRSS 420
QY 42 - - - - - A - - - - - - - - - - L - - - - - VP - - - - - P - - - - - 47
Db 421 VSFSSITVPDAPKVTWPEKTPGYLQPGQIPCSVDSLLPFTLFSVRNGVTGLGVQYLKESA 480
QY 48 - - - - - A - - - - - L - - - - - I - - - - - S - - - - - IQ - - - - - KG - - - - - 56
```



300 A-----K-----Q-----Q-----303  
2699 AYAI PASLSWYKGGPLKSDHVNIAANGHTLQI KEAQISDTRGYTCVASNIAGEDEL D 2758  
304 F-----P-F-----H-307  
2759 FDVNIQVPPSFOKLWEIGNMLDTCRNGEAKDVII NNPI SLYCETNAAPPPTLTWYKDGHP 2818  
308 --SA-----P-----A-----LDV-----D 315  
2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS D 2878  
316 -----WQ-----SN-----319  
2879 LPEEVTLVNKSALIECLSSGSPAPRNSWKDQGP LLEDHDKFSLNGRILQILNTQIT D 2938  
320 -----NT-----PAS--C-----S 326  
2939 IGRYVCVAENTAGSAKKYFNLNVHVPSPVIGPKSENLT VVVNNFISLTCEVSGFPDLS 2998  
327 -----T-----D-----MCI-----331  
2999 WLKNQOPIKLTNTLIVPGRTLQIIRAKVSDGEYTCI AINXAGESKKFSLTVVPPS 3058  
332 --H-----V--C-----K-----L--GO-----338  
3059 IKDHSLSLVNVRBGTSVSLECESNAVPPPVITWYKNGRMITESTHVEILLADGQMLHI 3118  
339 -----D-----R-----P-IK-----343  
3119 KKAESDGTQYVCRAINVAGRDKNFHLNVVPPSIEGP ERIVETISNPVLTCDATG 3178  
344 -----T-----F-----Q-----G-----H-----T-----349  
3179 IPPPTIAWLKNYKRIENSLSLEVRILSGSKLQIARSOHSDSGNYTCIASNMEGKAQKY 3238  
350 -----N-E-V-NA-----IKW--D--P--T-----360  
3239 FLSIQVPPSVAGAEIPSDSVLGENVELVCNANGIPTPLQIWLKDKGP IASGETERIV 3298  
361 --G--NLL-A--S-----C--S-----D-----MTL--373  
3299 SANGSTLNIYGALTS DTKYTCVATNPAGEEDRI FNLNVVYPTIRGNKDEAKELMTYD 3358  
374 -----K-----I-W-----SM-----KQ--D--NCV--384  
3359 TSINIECRXTGTPPPQINMLKNGLPLSLSHIRLAAGQVIRIVRAQVSDVAVYTCVASN 3418  
385 -----H-----D-----L-----Q-----388  
3419 RAGVDKNHYNLQVAFPNMDSMGTEBITVLKGSSTSMACITDGTAPSMAWLRDQPLG 3478  
389 -----Q-----ANLM-LAS-A--S--F-----KE 393  
3479 LDAHLTVSTHGMVLQLLKAETDSGKYTCIASNEAGEVSKHFKLVLEPPHNGSEHEER 3538  
394 I-----YT-----I-K-W-----399  
3539 ISVIVNPLELTCIASGIPAPKMTWMDGRPLPQTDQVQLGGGEVLRI STAQVEDTRY 3598  
400 -----SPTG-----P--GT-----N-----407  
3599 TCLASSPAGDDKKEYLVRVHVPVNIAGTDEPRDITVLRNQVLTCKSDAVPPPVITWLR 3658  
408 N-----P-----N-----ANLM-LAS-A--S--F-----420  
3659 NGERLQATPRVILSGRYLIQINNADLGDTANYTCVASNIAGTKTREFILTVNVPNIKG 3718  
421 -----DSTV-----RL-W--D--V-----DR-----G--I-----433  
3719 GPQSVILLNKSTVLECIABEGVPTPRITWTKDGA VLGNHARYSILENGFLHQSAHVTD 3778

434 --C-----I-----H-----TL--T-----439  
3779 TGRYLCWATNAAGTDRRRIDLVHVPPSIAPGPTNMTVI VNVQTTLACEATGPKPSINW 3838  
440 -K--H--O-E-----P-V--Y-----446  
3839 RKNGHLLANDQNSYRLSSGLSVIISPSVDDTATYECTVTNGAGDKRTVBLTVQVPP 3898  
447 SVA-----F-S-----P-D-----453  
3899 SIADEPTFLVTKHAPAVITCTASGVPPFSIHWTKNIGIRLLPRGDGYRILSSGAIEIAT 3958  
454 -----GRY-----L--ASG--S--461  
3959 QLNHAGRYTCVARNAAGSAHRHVTLVHBPVVIQOPSELHVILNPNPILLPCEATGTPSP 4018  
462 F-----D-----KCV-----466  
4019 FITWQKEGINVNTSGRNHVLPSGLQISRAVREDAGTYMCVAQNPA GTALGKILNVQV 4078  
467 --H-----I-W-----469  
4079 PPVISHLKEYIVAVDKPITLSCADGLPPPDI THWKGRAIVESIRQVRVSSGSLQIAF 4138  
470 -----N-----T-----Q-----472  
4139 VQPGDAGHYTCMAANVAGSSSTSTKLT VHVPPRI RSTEGHYTVNENSOAILPCVADGIPT 4198  
473 -----T-----G-----A-----L-VH 478  
4199 PAINWKONVLLANLLGKYTAEPYGEILILENVVLEDSGYFTCVANNAAGEDTHTVSLTVH 4258  
479 -----S-----Y-----R 481  
4259 VLPTTFELPGDVLNKGEOQLRUSCKATGIP LKLTWTFNNIIPAHFDSVNGHSELVIER 4318  
482 --GT-----G--G-----IF-----E--V--C-----490  
4319 VSKEDSGTYVCTAENS VGFKATGTVYVKEPPVFKG DYPSNMWIEPLGGNAILCEVKGDP 4378  
491 --WN-----A-----AGD-----496  
4379 TPTIQNRKGVDI EISHRIQLNGSLAIYGT VNEADAGDYTCVATNEAGVVERMSLTQ 4438  
497 -----K-V--G-----A-----SAS-----503  
4439 SPPIITLEPVETVINAGGKIILNCOATGEPOPTITW SROGHSISWDDRVNVLNNSLYIA 4498  
504 --D-----D-----GSV-----C-V-----509  
4499 DAQKEDTSBECVARNLMGSLVLRVPVIVQVHGFQSWANRACSVTCGKGIOKRSRLCN 4558  
510 --L-----DL--R-----K 514  
4559 QPLPANGKPKCGSDLEMCNQNK 4582

## RESULT 9

US-10-464-368-69  
; Sequence 69, Application US/10464368  
; Publication No. US20040023356A1  
; GENERAL INFORMATION:  
; APPLICANT: Krumlauf, Robb  
; APPLICANT: Ellies, Debra  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION  
; FILE REFERENCE: 40716-IP-017  
; CURRENT APPLICATION NUMBER: US/10/464,368  
; CURRENT FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/388,970  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 140  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 69

```

; LENGTH: 4599
; TYPE: PRT
; ORGANISM: MOUSE
US-10-464-368-69

Query Match      71.3%; Score 2610.5; DB 15; Length 4599;
Best Local Similarity 9.5%; Pred. No. 6.3e-30;
Matches 434; Conservative 56; Mismatches 16; Indels 4085; Gaps 365;

Qy 1 MS-----I-----S-----4
Db 1 MSQALLAILTSLGLPNAEVLIVGANQDHLCDPGEFLCHDHVTVCVSQMLCDGDPDPCPD 60
Qy 5 SD-----EV-----N-----F-----10
Db 61 QSDSLDTCPEVEIKPLNHIACHGSSACVHLSKLGNGVVDCPDGDFEGHCOELLPS 120
Qy 11 -L-----V-----Y-----R-----Y-L-Q-----17
Db 121 QQLNCQFKCAWRNATRCYCEDGEVAEDGRSKDQDECSYIGICSQTKNTYGSYACSC 180
Qy 18 -ESGF-----S-----H-----SAF-T-----F-----GIK-----S-----H-----33
Db 181 VE-GYIMQSDNRSCVKHEPTDKAPMLLISSLETIELFYINGSKWTTLSSANRNEIHTLD 239
Qy 34 -----I-S-QS-----NI--39
Db 240 FIYSEMICWIESRESSNQLKCGOITKAGRI TDQRIINSQSFQNVQEQMAFDMLTRNIYF 299
Qy 40 -----NG-----A-L-----VP-----45
Db 300 VDHVSDRIFVCFNFGSCVTLIESELENPKAIAADPIAGKLFFTDYGNVPKIERCDDGM 359
Qy 46 -----P-A-AL-----I-----SIIQKGLQ-----Y-59
Db 360 NRTRIVYSKAEQPSALALDLNRLVYVWLDLYDVGVVDYQGNRHTIVQ-GRQVRHLYG 418
Qy 60 -V-E-----AEVS-----IN-----I-----KESARGIRTYQKRTQVRSHACEV 477
Db 419 ITVFEDLYIA-TSSDNFIIRINFPNGTDIHSIIKESARGIRTYQKRTQVRSHACEV 477
Qy 69 D-----G-----TL-----F-----DGR-----76
Db 478 DAYMPGCGSHICLLSSSYKTRTCRCRTGFNMGSDGRSKRPNKELFLFYGKGRPGIVRG 537
Qy 77 -----PIE-----SL-----S-LI-----84
Db 538 MDLNTKIADECMIPENLVNPRALDPHAEANYIYFADTTSLFGRQKIDGTERETILKOD 597
Qy 85 D-----AV-----I-----M-----P-----89
Db 598 LDNVEGIADVWIGNNLTWTDNGHRTINVARLEKASOSRKTILEGGMHPRAIVDPVNG 657
Qy 90 -----D-V-----V-Q-----93
Db 658 WMYWTWKEDKIDDSVGRIEKAWMDGVNRQVFTSKMLWPENGLTLDFTSTLYWCDAAYD 717
Qy 94 -----T-R-----Q-----96
Db 718 HIEKVLNGTHRKVYSGKELNHPFGLSHGNVYVFTDYMGNSIFQLDLMTNEVTLRHE 777
Qy 97 -----QAY-R-----D-K-----LA-----104
Db 778 RAPLFGLOIYDPRKQGDNMCRINNGCGGLTCLIAIPAGRVACADNQLLDENGTTCTFNP 837
Qy 105 -----Q-----Q-----Q-----Q-----107
Db 838 EEIFHICKPGEFRCKNKHCIQARWKCDGDDCLDGSDESVTCFNHSCPDQFKQNNR 897
Qy 108 -----A-----AA-----AA-----AA-----116
Db 898 CIPKRWLCDGANDCGSNEBDSNOTCTARTCOAQDFSCNGRCIPTAWLDCRDDDCGDQTD 957
```

```

117 -AAS-----Q-----O-----121
958 EVASCEPPTCEPLTQFICKSGRCISNKHWCDDTDDCGDRSDRVCVHSCLDQFRCSSGR 1017
122 ---GS-A---KN---G---E---N-T---A---N-----GE-134
1018 CIPGWACDGDNDGDFSDETHINCTKEARSFAGCIGNFQCRPGDNCIPDLWRCDGK 1077
135 ---E---NGA---H---T---I---141
1078 DCEDGDEKGCNGTIRLCDHKTKFSCRSTGRCINNAWVCDGDVDCEDQSDDEDCDSCFLG 1137
142 ---AN-----NH-----145
1138 PPKYPCANDTSCVLOPEKLCNCRKDCPDGSDGDEGLDCECLNNGGCSNHCSSVVGRIVC 1197
146 ---T---D---MM-----E---VD-152
1198 SCPEGHQLKDNRTCEIVDYCASHLRCSQVCEQKHVMKSCVCEGALGTGDSCTSDS 1257
153 -----GD-----154
1258 FEAFIISIRHEIRRIDLHKGDYSLVPLRLNTIALDPHFNOQLLYWTDVVEDRIYRGKL 1317
155 -----VE-----I---157
1318 SESGGVSAIEVVVEHGLATPEGLTVDWIAGNIYDSDNLDQIEVSKLDGSLRATLIAGAM 1377
158 --P---P---N---K---161
1378 EHPRAIADPRYGILFTWDANDANFPRIESASMSGAGKRTYKDMKTGAWPGLTVDHFER 1437
162 -----AV-----VLRGHE-----SEV-----172
1438 RIVMTDARSDAIYSAFYDGTNMTIEIRGHEYLHPFAVSLYGEVYVTDWRTNTLAKANK 1497
173 -----F-----I---CA--176
1498 WTCQNVSVIQTSAQFDDLIQYHPSROPQPNPCANEGGPCSHCLINHNHSAACACP 1557
177 -----W-----NP-----VS---D---182
1558 HLMKLSDDKTKCYEMKFLLYARRSEIRGVIDNPNVNFITAFPTVIDDVAVIDPDASE 1617
183 --LL-----A---SG-----S-G---D---S---T---192
1618 ERLYWTDIKTITRAFINGTGLTETVISRDIOSIRGLAVDWVWSRNLVWISSEFDETOINV 1677
193 AR-----I-----W-----N-----197
1678 ARLDGSLKTSIIHGIDKPOCLAAHPVRGKLYWTDGNTINMANMDGNSKILFQKQKEPVG 1737
198 LS---EN---S---T---SG-----S-----T-----207
1738 LSIIDYVENKLYWISSNGTINRCNLDDGNLEVIEMKEELTRATALTIMDKKLWADQNL 1797
208 -OL-----VLR-----H-----213
1798 AOLGTCKNRDGRNPSILNRKTSVGHVMKYVDKEAQOGSNCQVNNGGCSQLCLPTSETTR 1857
214 -----C-----IR-----216
1858 TCMCTVGYVYLOKRMSCOGIESFLMSVHEGIRGIPLEPRDKVDALMPIGSAFAVGIDF 1917
217 -----EG-----218
1918 HAENDTIYWTMGLNKISRAKRDQWKEVDVNTGLRVEGIAVDWIAGNIYWTDHGFLNI 1977
219 -----GO-----D---V 222
1978 EVARLNGSPRYVLIISGLDQOPRSIAVHPKGLFTWTEWQVPCIGKARLDGSEKVMIVSV 2037
223 ---P---S---NK-----D---V---T---S-LD-----232
```

Db	2038	GITWPNGISIDYEENKLYWCDAKSDKIERIDLTGCRNREVLSSGNDLFSVAVFGAYII	2097	Db	3114	SKLNGLYPTLVSKRLKFPRLDLSLDPFRAGNLYWIDCCBYPHIGRVMGMDGTNOSVVIETKI	3173
Qy	233	W-----N-S-----	238	Qy	368	S-----D-----D-----M-----	376
Db	2098	WSDRAHANGSVRRGHKNQDATTETVMTGTGLGVNLKEIKIFNRVREKTNVCAKENGCCQOL	2157	Db	3174	SRPMALTIDYVNHRLYWADEHNHIEFSNMDGSHRHKVPNQODIPGVIALTLFEDYIYWDGK	3233
Qy	239	-----LLA-----T-----G-----	245	Qy	377	-----S-MK-----Q-D-----N-----C-----	383
Db	2158	CLYRGNRRTCAAGYLAGDGTCLRHGGLYSGRTILKSIHLSDETNLNSPVRPYEN	2217	Db	3234	TKSLSRVHKTSGADRLSLNSHWAITDIQVHVSYPQDVSKHLCTVNNGCCSHLCLLPG	3293
Qy	246	-----D-----G-----F-----A-----R-----I-----	251	Qy	384	VH-----D-----L-----	387
Db	2218	PNYFKNIIALAFDYNQRREGTNRIFYSDAHFGNIOLIKONWEDROVIVENVGSGVEGLAYH	2277	Db	3294	KTHTCACPTNFYLAADNRNCTLSNCTASOFRCKTDKICIPFWKCDTVDDCGDSDDEDDCP	3353
Qy	252	-----W-----TK-----D-----G-----	258	Qy	388	-----Q-----Q-----H-----NKE-I-----	394
Db	2278	RAWDLYTSSSTSSITRHTVDTQTPGAIDREAVITMSDDHPHVALDECONLFWTWM	2337	Db	3354	EFKCPGRFQCCTGLCALPAFICDGENDCGNSDELCDTHVCLAQCFKTKNKKCIPVN	3413
Qy	259	-----A-----ST-----	264	Qy	395	-----Y-----T-----I-K-W-----SP-----	401
Db	2338	NEQHPSIMBATLTGKNAHVVSVDILTTPNGLTIDHRAEKLIFSGLGKIERCEYDGSOR	2397	Db	3414	LRCNGQDCGDEDEKDCPENSCTSPDYFOCKTTKHICISKLWVCDDEDDCADASDEANCDK	3473
Qy	265	H-K-GP-----IP-ALKW-----NK-----K-----	277	Qy	402	T-GP-----G-----T-N-----N-P-----	409
Db	2398	HVIVSGPGTFLSLAVYDSYIFWS-DWGRRAILRSNKYTGGETKILRSIDPHQPMGIITAV	2456	Db	3474	KTCGPHEFOCKNNNCIPDHWRCDNQDCSDSDENCKPQTCTLKDFLCNSGDCVSSRFW	3533
Qy	278	-----I-----W-----D-----D-----AH-----	296	Qy	410	-----N-----A-----N-----	412
Db	2457	ANDTNSCELSPCALLNGGCHDLCLLTPDGRVNCRCGRVLLANNRCVTNNSCNIYSEF	2516	Db	3534	CDGEFDCADGDEKNCETSCSKDQFCNSGQCLSAKWKCDGHEDCKYGEDEKNCBPAPV	3593
Qy	281	-----I-----L-----S-AG-----V-----	286	Qy	413	-----LMLA-----SAS-----F-----	420
Db	2517	ECNGDCVDYVLTCDGIPHCKDSDKLLYCNRSRSGFKPCYNRRCVPHGKLCGDTND	2576	Db	3594	CSSESVMCASGCLSLKNGEPCVDSGDEMDCVIECKEDQFOCKNKAYCIPIRWLCD	3653
Qy	287	-----D-K-----T-----T-I-----	291	Qy	421	-----D-S-----T-----V-R-----L-W-----D-V-----	429
Db	2577	CGDSSDELCKVSTCTVFRCADGTCPISARCQNQMDCSDASDEKGCNNTDCTHFYKL	2636	Db	3654	GIYDCVDSDEETCGRGGISICRDDEFLCNLSLKLHFWDCGEDDCGDSDEAPDMCVKF	3713
Qy	292	-----I-----W-----D-----D-----AH-----	296	Qy	430	-----DR-----GI-----	433
Db	2637	GVKSTGPIRCNSTSLCVLPSWICDGSNDGCDYSDCLKPCVQNKHCENYFCGPGRCIL	2696	Db	3714	LCPPTPRYRCNRDRIQLQLEKICNGINDCGDSDSEHSCGKLSKSKPKCKDEFTCSNRN	3773
Qy	297	-----G-E-----A-----K-----	301	Qy	434	CI-----HT-----	437
Db	2697	NTWVCGQKDCEDGLDELHCDSSCNWQFACSVKCKISKHWICDGEDDCGDSLDESISIC	2756	Db	3774	CIPMELQDCLDDCGDSDGDEQCLKTPTEHTCENNGNPGDDAYCQNKIKTSVFCRCKPGF	3833
Qy	302	-----Q-----Q-----F-----	304	Qy	438	-----L-T-----K-----	440
Db	2757	GAVTCAADMFCQGGSHACVPOHWCDEGRDCPDGSDDELSSAGCAPNNTCDENAFMCHNKV	2816	Db	3834	QRNMKGRECADLINECLLFGICSHHCLNTRGSYKVCVQDNQFQKNNNSCIAKGSBDQALYIA	3893
Qy	305	-----P-P-H-----S-----AP-----ALD-----VD-----WQ-----	318	Qy	441	-----HQ-----E-----P-V-Y-----	446
Db	2817	CIPKQFVCDHDDCGDSDGDEFLQCYRQGPBEFPCA-DGRCLVNTLMQCDGDFPCPDSS	2875	Db	3894	NDTDILGFVPYPNYGGHQIISHVEHNSRITGMDVHYQBNVILWSTQFNPFGIFYKNIDA	3953
Qy	319	-----N-----NT-FASC-----STD-MC-I-----HV-C-----	334	Qy	447	-----S-----VA-----FS-----	451
Db	2876	DEAPINPRCSAEHSCNSFFM-CXNGRCIPS-DGLCDIRDDCGDSDGDETNCHINECLSK	2933	Db	3954	REKQANSGLICPEKPRPRDIAVDWAGVNYWTDHSMHWFYSYTTHTWTSRLYSINVGOL	4013
Qy	335	KL-G-QD-R-PI-----KT-----F-----Q-----G-----	347	Qy	452	-----P-----DG-----	454
Db	2934	KISGCSQCDLPSVYKCKWPGFQKDDGKTCVDIDECSSGFPSCQICNTYGTVKCHC	2993	Db	4014	NGPNCTRLLTNMBEPIAYAVNPKRGMWYTVIGDHSHEEAAAMDGTLLRRVLVQKNLQRP	4073
Qy	348	-----H-----H-----T-----N-----	350	Qy	455	-----R-Y-----LA-S-----GS-----	461
Db	2994	ABGYETQDPNPGCRSLSDPEPFLILADQHQRKISTDGSNTYLLKQGLNNVIALDFDYR	3053	Db	4074	TGLTVDFGERIYADFELSIGSVLYDGSPPVSVSSKQLLHHRIDVDFEYIYGAP	4133
Qy	351	-----E-----V-NA-----I-K-----W-D-----	358	Qy	462	-----F-----DK-----	464
Db	3054	EEFIWIDSSRPNGSRINRCNLGSDIKVHTATVPNALAVDWIGKLYWSDEKRIIEV	3113	Db	4134	KNGIFRVQKFGHGSVEVLALGVDTKTSILVSHRYKQLNLPNCLDSCDFLLNPSGAT	4193
Qy	359	-----PT-----GNLL-ASC-----	367	Qy	465	CV-----H-----I-----W-----N-----	470
				Db	4194	CICPEGKYMNGTCHDDSLDDCKLTCEGRCILNEKGDILRCHWPSYSGRCVNHNC	4253

QY 471 T---Q---T-G---AL---V---H---S--- 479  
Db 4254 SNYQNGGTCTPSTLGRPTICALGFTGPNCGKAVCEDSCHNGGSCVVTAGNOPYCHCOA 4313  
QY 480 Y---R---G---T--- 485  
Db 4314 DYTGRQYVYCHYCVNSBSCTIGNDGSVECVPTRYEGPKCEIDKVCRCGHGCIINK 4373  
QY 486 ---IF---E---VC---W--- 491  
Db 4374 DNEDIFCNCTNGKIASSCQLCDGYCYNGGTCLQDPETSPFVVCVSTNWSGTQCPAPKS 4433  
QY 492 --- 491  
Db 4434 SKSEHISTRSAIIVPLVLLVTLVTLVIGLVCKRRTKTRRQPIINGGINVEIGNP 4493  
QY 492 --NA---AG---D-K---VG---A---SA---SD---G 505  
Db 4494 SYNMYEVDHDSGGLLEPSFMIDPVKSRVYIGGSSAFKLPHTAPPIYLNLSDLKGLPTFG 4553  
QY 506 ---S-V---C---VLDLRK 514  
Db 4554 PTNYSNPVYAKLYMDGQNCNRSLASV-DERK 4583

RESULT 10  
US-10-464-368-70  
; Sequence 70, Application US/10464368  
; Publication No. US20040023356A1  
; GENERAL INFORMATION:  
; APPLICANT: Krumlauf, Robb  
; APPLICANT: Ellies, Debra  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION  
; FILE REFERENCE: 40716-IP-017  
; CURRENT APPLICATION NUMBER: US/10/464,368  
; CURRENT FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/388,970  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 140  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 70  
; LENGTH: 4599  
; TYPE: PRT  
; ORGANISM: MOUSE  
US-10-464-368-70

Query Match 71.3%; Score 2610.5; DB 15; Length 4599;  
Best Local Similarity 9.5%; Pred. No. 6.3e-30;  
Matches 434; Conservative 56; Mismatches 16; Indels 4085; Gaps 365;

QY 1 MS-----I-----S----- 4  
Db 1 MSQLLAILTLGELLENAEVLIVGANQDHLCDPGEFLCHDHVTCVWSQWLCDGDPDPCPD 60  
QY 5 -SD-----EV---N-----F----- 10  
Db 61 QSDLESIDTCPEVEIKPLNHIACHGSSACVHLSKLCNGVVDPCDFDEGHCHQELLPSCL 120  
QY 11 -L-----V-----Y-----R-----Y-L--Q----- 17  
Db 121 QQLNCQFKAMVRNATRCYCEGFEVAEDGRSCQDQECISYICSCQTKNTYGVYACSC 180  
QY 18 -ESGF-----S-----H-----SAF-T---F---GIK-----S-----H--- 33  
Db 181 VE-GYIMQSDNRCKVKHEPTDKAPMLLISSLETIELFYNGSKMTTLLSANRNEIHTLD 239  
QY 34 -----I-S-QS-----NI--- 39  
Db 240 FIYSEEMICWIESRESSNLKCCQITKAGRLTDQRIINSLSQFQNVQMAFDWLTNIYF 299  
QY 40 -----NG-----A-L-----VP----- 45

Db 300 VDHVSDRIFCVFNENGVCVTLIESELHNPKAIAADPIAGKLFPTDYGWVPKIBERCLDGM 359  
QY 46 -----P-A-AL-----I-----SITQKGLQ---Y- 59  
Db 360 NRTRIIVSKAEOQSALALDLVNRVYVLDYLYDVGVVDYQGRNRHTIVQ-GRQVRLHYG 418  
QY 60 --V-E---AEVS---IN-----E- 68  
Db 419 ITVPEDYLYA-TSSDNFNIIRNFNGTDIHSIIKMSARGIRTYQKRTQPTVRSACEV 477  
QY 69 D-----G-----TL---F---DGR----- 76  
Db 478 DAYMEGCGSHICLLSSSYKTRICRTRCTGFNMGSDGRCKRPNKELFLFYKGRPGIVRG 537  
QY 77 -----PIE---SL-----S-LI----- 84  
Db 538 MDLNTKIADECMIPENLVNPRALDPHABANYIYFADTTTSLIGROKIDGTERTILKOD 597  
QY 85 -D---AV-----M-----P--- 89  
Db 598 LDNVEGIAVDWIGNNLYWTNDGHRKTINVARLEKASQSRKTLLEGNSHPRAIVDPVNG 657  
QY 90 -----D-V-----V-Q----- 93  
Db 658 WMYTWDKEDKIDDSVGRTEKAMWDGVRQVFTSKMLWPNGLTLDFTHTSTLYWCDAVYD 717  
QY 94 -----T-R-----Q----- 96  
Db 718 HIEKFLNGTHRKVVYSGKELNHPFGLSHHGYVFWTYDYMNGSIFOLDLMTNVTLLRHE 777  
QY 97 -----QAY--R---D-K-----LA----- 104  
Db 778 RAPLGLQIYDRKQOGDNMCNINNGCGTLCIAIPAGRVACADNQLLDENGTTCTFNP 837  
QY 105 -----Q-----Q-----Q-----Q----- 107  
Db 838 EBIRFHICKPGEFRCKNKHCIQARWKCDGDDCLDGSDEDSVTCFNHSCPDQFKQNNR 897  
QY 108 -----A-----AA---AA---AA---AA---AA- 116  
Db 898 CIPKRWLCDGANDCGSNEDESNOTCTARTCOAQDFSCGNRCIPTAWLCDREDDCGDQTD 957  
QY 117 -AAS-----Q-----Q-----Q----- 121  
Db 958 EVASCEPTCEPLTQICKSGRCISNKHCHDITDDCGDRSDEVGCVHSCLDLQDFRCSSGR 1017  
QY 122 ---GS-A---KN-G---E-N-T---A-----N-----GE- 134  
Db 1018 CIPGHWACDGDNDGDFSDETHINCTKEARSPAGCIGNEFQCRPDGNCIPDLWRCDGEK 1077  
QY 135 -----E---NGA---H---T---I----- 141  
Db 1078 DCEGSDGKGCNGTIRLCHDKTKFCSRGTGRGINNAWVCDGVDCEQSDSEBDCSFLCG 1137  
QY 142 ---AN-----NH----- 145  
Db 1138 PPYPFANDTSVLOPEKLCNCRKDCPDGSDGDLCDCECLNNGGCSNHCSSVVPGRGIYC 1197  
QY 146 -----T---D-----MM-----E-----VD- 152  
Db 1198 SCPEGHQLKKNRTCEIVDYCAHLRCSQVCRQKHVMKCSYEGWALGTDGESCTSDS 1257  
QY 153 -----GD----- 154  
Db 1258 FEAFIIFSIHRIDILHKGDYSLLVPLGRNTIALDPHFNSLLYWTVDVEDRIVRGKL 1317  
QY 155 -----VE-----I----- 157  
Db 1318 SSSGGVSAIEVVVVEHGLATPEGLTVDWIAGNIYWIDSLNDQIEVSKLDGSLRATLIAGAM 1377  
QY 158 --P-----P-----N-----K----- 161  
Db 1378 EHPRAILDPRYGLFWTDWDANFPRIEASMSGAGRKTIYKDMKTGAWPNGLTVDFHFER 1437



QY 162 -----AV-----VLRGHE-----SEV----- 172  
Db 1438 RIVMTDARSDAIYSAFYDGTNMIIEIRGHEVLSHPFAVSLYSEVYWTWRTWLAKANK 1497  
QY 173 -----F-----I-----CA-- 176  
Db 1498 WTCQNVSVIOKTSQAOPFDLQIYHPSRQPOAPNPCAANEGRGPCSHLCLINHRNSAACACP 1557  
QY 177 -----W-----NP-----VS--D-- 182  
Db 1558 HLMKLSDDKTCYEMKKFLYARREIRGVGDIDNPYNFITAFTVPDIDDAVIDPDASE 1617  
QY 183 --LL--A--SG--S--G--D--S--T-- 192  
Db 1618 ERLYWTDIKTQITRAFINGTGLTVISRDIQISIRGLAVDWVRNLYWISSEFDETOINV 1677  
QY 193 AR-----I-----W-----N----- 197  
Db 1678 ARLDGSLKTSIIHGIDKPOCLAAHPVRGKLYWTDGNTINMANMDGNSKILFQNKPEVG 1737  
QY 198 LS--EN--S--T--SG--S--T-- 207  
Db 1738 LSIYVENKLYWISSNGTINRNLGGNLEVIEMKEELTKATALTIMDKKLWADQNL 1797  
QY 208 -QL--VLR--H----- 213  
Db 1798 AOLGTCKRDRNPISILRNKTSGVHMKVYDKEAQOQSGNSCVNNGGCSQLCLPTSETTR 1857  
QY 214 -----C-----IR----- 216  
Db 1858 TCMCTVGYVLQKNRMSQGIIESFLMYSVHGEIRGIPLEPRDKVDLMPISGAFAVGIDF 1917  
QY 217 -----EG----- 218  
Db 1918 HAENDTYWTDMLGINKISRAKRDQTKBEDVVTNGLGRVEGIAVDVIAWTDHGFNLI 1977  
QY 219 -----GO-----D-----V 222  
Db 1978 EVARLNGSPRYVLIISQGLDQPSIAVHPEKGLFWTEWGOVPCIGKARLDGSEKWLVSU 2037  
QY 223 -P--S--NK--D--V--T--S--LD-- 232  
Db 2038 GITWPGISIDYBENKLYWCARSDKIERIDLTGANREVLVSGSNVDLFSVAVFGAYIY 2097  
QY 233 W-----N--S-----E--CT-- 238  
Db 2098 WSDRAHANGSVRRGHKNDAETVTMTRTGLGNLKEIKIFNRVREKGTNVCAKENGCCOQL 2157  
QY 239 -----LLA--T-----G-----S--Y-- 245  
Db 2158 CLYGRNSRRTCAAGHYLAGDGVTCRLRHEGYLLYSGRITLKSITHLSDETNLNSPVRPYEN 2217  
QY 246 -----D--G--F--A-----R--I----- 251  
Db 2218 PNYFKNIIALAFDYNORREGTNRIFYSDAHFGNIQLIKDNWEDROQVIVENGVSGEGLAYH 2277  
QY 252 -----W-----TK--D--G-----NL-- 258  
Db 2278 RAMDTLYWTSSTSTSTRHTVDTQTRFCAIDREAVITMSEDDHPHVLALDECQNLMEFTNW 2337  
QY 259 -----A--ST-----LG-----Q-- 264  
Db 2338 NEQHPSIMRATLTGKNAHVWVSDIITPGLTIDHRAEKLYFSDGSLGKIERCEYDGSQR 2397  
QY 265 H--K--GP-----IF--ALKW-----NK-----K----- 277  
Db 2398 HVIVKSGPGTFLSLAVTDSYIFWS--DWGRRAILRSNKYTGTETKILRSIDIPHQPMGIIV 2456  
QY 278 -----G-----N-----F 280  
Db 2457 ANDTNSCELSPCALLGGCHDLCLLTPDGRVNSCRGDRVLLANNRCVTNKSNCITYSEBF 2516

QY 281 -----I-----L-----S--AG-----V----- 286  
Db 2517 ECGNGDCVDVYLTCDDGI|PHCKDKSDEKLLYCNRSRSGFKPCVNRRCVPHGKLCDCGTND 2576  
QY 287 -----D--K-----T-----I----- 291  
Db 2577 CGDSSDELCKVSTCSTVEFRCADGTCIPRSARCNQNMDCSDASDEKGCNNTDCTHFKYL 2636  
QY 292 -----I-----W-----D-----AH----- 296  
Db 2637 GVKSTGFIRCNSTSLCVLPSWICDGSNDGSDYDELKCPVQNKHKKEENYFGCPSGRCIL 2696  
QY 297 -T-----G--E-----A-----K----- 301  
Db 2697 NTWVCDGXDCEDGLDELHCDSSCNQFACSVKCKISKHWICDGEDDCGSDLESDSIC 2756  
QY 302 -----Q-----Q-----F----- 304  
Db 2757 GAVTCAADMFSQGSFHACVPHWLCDBGERDCPDGSDLSAGCAPNNTCDENAFMCHNKV 2816  
QY 305 --P--F--H-----S-----AP-----ALD--VD--WQ-----S 318  
Db 2817 CIPKQVCDHDDCGDGSDEFLOCGYRQCGPEFRCA--DGRCLVNTLWQCDGDFCDPSS 2875  
QY 319 -----N-----NT--FASC-----STD--MC--I-----HV--C-- 334  
Db 2876 DEAPINPRCSAEHSCNSFFM--CKNGRCIPS--DGLCDIRDDCGDGSDETCHNECLSK 2933  
QY 335 KL--G--QD--R--PI-----KT-----F--Q-----G----- 347  
Db 2934 KISGSCQDQDLPSVYKCKWPGFQLKDGKTCVDIDECSSGFPSCQOCINTYGTVKCHC 2993  
QY 348 -----H-----T-----N----- 350  
Db 2994 AEGYETQPNPNCGRSLSDDEEPLIADQHEIRKISTDGSNYTLKQGLNVLALDPDYR 3053  
QY 351 -E-----V--NA-----I--K--W--D----- 358  
Db 3054 EEFYIWDSSRPNRSRINRMLNGSDIKVHVHTAVFNALAVDWIGKNLYWSDTEKRIIEV 3113  
QY 359 -----PT-----GNLL--ASC----- 367  
Db 3114 SKINGLYPTVLVSKRLKFPRLSLDPRAGNLYWIDCEYPHIGRVGMGTNQSJVETKI 3173  
QY 368 S-----D-----M-----TL--K--I--W----- 376  
Db 3174 SRPMALTIDYVNRHLYWADENHIEFNSMDGSHRHKVPNOIPGVIALTLFEDYIYTDGK 3233  
QY 377 -----S--MK-----Q--D-----N-----C----- 383  
Db 3234 TKLSRVHKTSGADRLSLNSMHAITDIOVYHSYRQPDYVSKHLCTVNNGGCSHLCLLPGP 3293  
QY 384 -VH-----D-----L----- 387  
Db 3294 KTHTCACPTNFYLAADRNTCLSNCTASQFRCKTDKCIPEFWMKDVTDDCGDGSDEPDDCP 3353  
QY 388 -----Q-----Q-----H-----NKE--I-- 394  
Db 3354 EFKCQGRFCGTGLCALPAFICDGENDCGNSDELNCOTHVCLAGQFKCTKNKCLIPVN 3413  
QY 395 -----Y-----T-----I--K--W-----SP----- 401  
Db 3414 LRCNGODDCGDEDEKDCPENSCSPDYFOCKYTKHCISKLVWCEDEPDCADEANCDK 3473  
QY 402 -T--GP-----G-----T--N-----N--P----- 409  
Db 3474 KTCGPHEFOCKNNNCIPDHWRCDNQDSDNSDEDNCKPQTC|LKDPLCSNGDCVSRFW 3533  
QY 410 -----N-----N-----A-----N----- 412  
Db 3534 CDGEFDCADGSEKNCETSCSKDQFOCSNGQCLSAKWKCDGHEBCKYGEDEKNCPEFPV 3593  
QY 413 -----LMLA-----SAS-----F----- 420

```
Db 3594 CSSEYWCAGGCLASLCKNGEPDCVDGSDMDCVIECKEDQFCCKKAYCIPRWLCD 3653
Qy 421 -----D-S-----T-----V-R-----L-W-----D-V-- 429
Db 3654 GIYDCVDGSDDETCGRGSGICRDEFLCNLSLCKLHFVWCDGEDDCGNSDEAPDMCVKF 3713
Qy 430 -----DR-----GI----- 433
Db 3714 LCPTRPYRCNDRICIQLEKICNGINDGNSDEHCSGKLSLKSPCKKDEFTCSNRN 3773
Qy 434 CI-----HT----- 437
Db 3774 CIPMELQCDLDDCGSDGDEGCLKTPIHTCENNGNPGCDAYCNOIKTSVFCRCCKPGF 3833
Qy 438 -----L-T-----K----- 440
Db 3834 QRNMKGRECADNELLFGICSHHCLNTRGSKVCQDNFQEKNNSCIAKSGEDQALYIA 3893
Qy 441 -----HO-----E-----P-V-Y----- 446
Db 3894 NDTILGVFPFNYSGGHQQLSHVHNSRITGMDVHYQRNVIVWSTQFPGGIFYKMIDA 3953
Qy 447 -----S-----VA-----FS----- 451
Db 3954 REKQANSGLICPEFKRPRDIAVDWAGNVYWTVDHSMWFYSYTHWTSLRYSINVGQL 4013
Qy 452 -----P-----DG----- 454
Db 4014 NGPNCTRLTNMAGEYAIANPRKGRMYMTVIGDHSHEAAMDGTLRRLVQKNLQRP 4073
Qy 455 -----R-Y-----LA-----S-----GS----- 461
Db 4074 TGLTVHFGRIYADPELSIGSVLYDGSPPVSVSSKQGLLHPHRIDVFEDIYAGP 4133
Qy 462 -----F-----DK----- 464
Db 4134 KNGIFRVQKFGHGSVEVLALGVDTKSLVSHRYKQLNPNCLDLSCLFLLNPSGAT 4193
Qy 465 CV-----H-----I-----W-----N-- 470
Db 4194 CICEGYMMNGTCHDDSLDDSCKLCFCENGGRILNEKGLRCHCWPSYSGRCVYNHC 4253
Qy 471 T---Q-----T-G-----AL-----V-----H-----S----- 479
Db 4254 SNYCQNGTCTPILGPTICALGFTGPNCGKAVCEDSCHNGSCVVTAGNQPYCHCQA 4313
Qy 480 Y---R-----G-----T----- 485
Db 4314 DYTGRQYVCHRYCVNSBESCTIGNDGSVECVCPTRYEGPKCEIDKVCRCGHGCIINK 4373
Qy 486 ---IF-----E-----VC---W----- 491
Db 4374 DNEDIFCNCTNGKIASSCOLCDGYCYNGGTCLDPETSIPVCVSTNWSGTQCPAPAKS 4433
Qy 492 -----S-V-----C-----VLDLRK 514
Db 4554 PTNYSNPVYAKLYMDGQNCNRNSLASV-DERK 4583
```

RESULT 11

US-10-120-801-74

; Sequence 74, Application US/10120801

; Publication No. US20030203843A1

; GENERAL INFORMATION:

```
; APPLICANT: Pena, Carol
; APPLICANT: Guo, Xiaojia
; APPLICANT: Shimkets, Richard
; APPLICANT: Padigaru, Muralidhara
; APPLICANT: Kekuda, Ramesh
; APPLICANT: Spytek, Kimberly
; APPLICANT: Mehraban, Fuad
; APPLICANT: Topper, James N.
; APPLICANT: Malyankar, Uriel
; APPLICANT: Wasserman, Scott
; APPLICANT: Edinger, Shlomit
; APPLICANT: Smithson, Glennda
; APPLICANT: Gunther, Erik
; APPLICANT: Komuves, Laszlo
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same
; FILE REFERENCE: 21402-340
; CURRENT APPLICATION NUMBER: US/10/120,801
; CURRENT FILING DATE: 2002-04-11
; PRIOR APPLICATION NUMBER: 60/285748
; PRIOR FILING DATE: 2001-04-23
; PRIOR APPLICATION NUMBER: 60/286068
; PRIOR FILING DATE: 2001-04-24
; PRIOR APPLICATION NUMBER: 60/286292
; PRIOR FILING DATE: 2001-04-25
; PRIOR APPLICATION NUMBER: 60/288334
; PRIOR FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: 60/291241
; PRIOR FILING DATE: 2001-05-16
; PRIOR APPLICATION NUMBER: 60/322284
; PRIOR FILING DATE: 2001-09-14
; PRIOR APPLICATION NUMBER: 60/385609
; PRIOR FILING DATE: 2001-04-20
; NUMBER OF SEQ ID NOS: 155
; SOFTWARE: Patent in Ver. 2.1
; SEQ ID NO 74
; LENGTH: 5175
; TYPE: PRT
; ORGANISM: Caenorhabditis elegans
; US-10-120-801-74
```

Query Match 71.0%; Score 2596.6; DB 14; Length 5175;

Best Local Similarity 9.0%; Pred. No. 2.1e-29;

Matches 454; Conservative 48; Mismatches 10; Indels 4514; Gaps 390;

```
Qy 1 M-----SI-----SS-----D-----E--- 7
Db 1 MGRSPWLVGLGLLLATTCSNVDDKNDPTGKSLAFVFDITGSMFDDLVQVREGAAK 60
Qy 8 ---V-----N-----F---L---VY----- 13
Db 61 IFKTWQREKLIYNYIMVFPDPLGEIINTTDSYFMRQLSKVYVHGGDCPEKTLTG 120
Qy 14 -----R---Y-----L---QE---SG---F-----SHSAF-T 27
Db 121 ILKALQISLPSSPIYVFTDARSKDYLHLEDEVLNTIQEKSSVVFVMTGCGNRTHPGFT 180
Qy 28 -----FG-I---KS-----H-ISQ-----S-NI----- 39
Db 181 YEKIAAASFGQVPHLEKSDVSTVLEYVRHAVKOKVHLMYERAGGTVSRNIPVDKHL 240
Qy 40 -----N-----GA-----LV---P---P----- 46
Db 241 ELTISLGGDKDSDNLDIVLDPDPEGTVDKRLYSKSGGITDKNVKILRLKDPSPGVWTV 300
Qy 47 -----A-----A-----LI---S---I----- 52
Db 301 NTSRLKHTIRVFGHGAVDKYGFAFRPLDRIELARPRVFNODTYLLINMTGLIPPGTV 360
Qy 53 --I-----Q-----KGL-----Q-Y---VE---AE- 63
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPPVPKGLFFVRVQGVDEDNEMFIAPT 420
Qy 64 -----V-----S-I-----N-----ED---GTLF----- 73
```

Db	421	AIGSVVGGPRAPFMSPIHQEFVGRDLNLCTVESASAYTIYWKGTGEDIIGGPLFYHNTD	480
Qy	74	-----D-----G-----R--P-I-----E-----	79
Db	481	TSVWTIPELSLOAGEYECRVISNNGNYSVKTRVETRESPEIFGVGVNVSPLGEAAFLH	540
Qy	80	-----S-----L-----SL-I-----DA-----	86
Db	541	CSTRSAGEIRWTRYGATVFNGBNTERNPTNGTLKIHVTRADAGVYECWARNAGCMST	600
Qy	87	-----VM-----P-DV-----:-----V-92	
Db	601	RKMRDLMEPPSVKVTQDVYFNMRBGVNLSCAMGDPKPEVHWYFKGRHLLNDYKYQVG	660
Qy	93	Q-----T-----R-Q-QAYRD-----KLA-----Q-----	105
Db	661	QDSKFLYRDATHHDEGTCECRAMSOAQAR-ROTTDLMLATPPKVEIIQNMWVGGRDRV	719
Qy	106	-----Q-----Q-----Q-----A-----A-A-----A-----	111
Db	720	SFEKTIIRGPKHPRFKNGKOLIKPDDYIKINEGQLHMGAKDEADAGAYSCVGENMAG	779
Qy	112	-----A-----A-----A-----A-----A-----A-----	114
Db	780	KDQVANLSVGRVPTIIESPHTVRVNIERTQVTLQCLAVGIPPPPEIEWQKGNVLLATLNNP	839
Qy	115	-----A-----A-----A-----ASQ-----	120
Db	840	RYTQLADGNLLITDAQIEDOGQFTCIARNTYGOOSQSTTLMTGLVSPVLGHVPPPEROLI	899
Qy	121	--Q-----Q-----GS-----	123
Db	900	EGQDLTLCVWVLGTGPKPSIWIKDDKPVEEGPTIKIEGGSLRLRGGNPKDEGKYTCI	959
Qy	124	-----A-----A-----K-----N-----GE-----	128
Db	960	AVSPAGNSTLHINVLKIKPEFVYKPEGGIVFKPTISGMDKEHAVVNVTHDVLGEGFA	1019
Qy	129	-----A-----NT-----A-N--G-133	
Db	1020	IPCVSGTTPPIITWYLDGRPITPNSRDFVTADNTLIVRKADKSYSGVTCQATNSAGD	1079
Qy	134	E-----E-----E-----E--N-136	
Db	1080	NEQKTIIRIMNTPMISPGQSSFNWVDDLFTIPCDVYGDPKPVTLLDDKPFTEGVVNE	1139
Qy	137	G-----AH-----TI-A-N-----N-----	144
Db	1140	DGSLTIPNVNEAHRGTFTCHAQNAAGNDTRVTLVHTTPTINAENOEKIALQNDIVLE	1199
Qy	145	-----HT-----D-----MME-----V-----D-152	
Db	1200	CPAKALPPPVRMLWYEGEKIDSQLIPHTIREDGALVQNKLENTGVFCQVSNLAGEDS	1259
Qy	153	-----G-DV-----BIP-----P-----NK-AV-----	163
Db	1260	LSYTLVHKPKIIEVPGVVDVVKGTIEIPCRATGVPEVIRTWKNGIDLKMKDEKFS	1319
Qy	164	V-----LR-----G-----H-----E-----S-----	170
Db	1320	VDNLGLTIRIYEADKNDIGNCYNVVTNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVGDR	1379
Qy	171	-----E-----V-F-----I-C-----	175
Db	1380	VELKCVVEASPPASVTWFRRGIAIGTDTKGVVVEDGTLVIQASVEDATIYTKASNPA	1439
Qy	176	-----A-----	176
Db	1440	GKAENLQVTVIASPDIKDPDVVTQBSIKESHPSFLYCPFSNPLPOISWYLNDKPLIDD	1499
Qy	177	--W-----N-----	178

[illegible]

QY 284 ---AG---V---DK---T---TI---I--- 292  
Db 2640 SGRYTCIVRNPAEARKLFDFAVNDPPSISDELSSANIQTIVPYVPEINCVVSGSPHPK 2699  
QY 293 ---W---D---A---H---T---G--- 298  
Db 2700 VYMLFDDKPLEPDSAAVELTNGETLKVRSQVEHAGTYTCEAQNNVGKARKOFLVRVTA 2759  
QY 299 ---E---AK---Q--- 302  
Db 2760 PPHEKEREYVARVGTMLLTCAEBSVPLSSVYHHADESQVGVITSKYAANEKTLN 2819  
QY 303 ---Q---F--- 304  
Db 2820 VTNILQDEGFYCTAVNEAGITKFPKLVITPYFLDQOKLYPIILGKRLTLDCSATG 2879  
QY 305 ---P---F--- 306  
Db 2880 TPTPTILFMKDGKRLNESDEVDIIGTLVIDNPQKEVEGRYTCIAENKAGRSEKDMVVEV 2939  
QY 307 ---H---S--- 308  
Db 2940 LLPKLSKEWINEVOAGDELTECPIEDTSGVHITWSROFGKDGQDLMRAQSSSKSL 2999  
QY 309 ---A-P-A-L---D---V-D--- 315  
Db 3000 YIMOATPEDADSYSIAVNDAGBAEVQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRT 3059  
QY 316 ---W---Q--- 317  
Db 3060 EGIPPPESISFLDGKPILEMPGVYKQGLSLRDNKPNQEGRYTCVAENKAGRABQDT 3119  
QY 318 ---S---N---T---F--- 322  
Db 3120 YVEISEPRVMASEVNRVVEGRQTTIRCEVFGNPEVFNWVNLKDGEPYTSLLQFSTKLS 3179  
QY 323 ---A-SC---S-TD---M-C-330 330  
Db 3180 YLHLRETTLDGGTYTCIATNKAGESQTTDVEVLVPPRIEDEERVLQKEGNTYWHQC 3239  
QY 331 ---I---HV---C---K 335  
Db 3240 VTGRPVVYTKRNGKEIEQFNPVLHNRATRADEKGYSCIASNEAGTAVADFLIDVFTK 3299  
QY 336 ---L-GO---D---RP---I-K-343 343  
Db 3300 PTFETHETTFNIVEGSAKIECKIDGHPKPTISWLKGRPFNMNIIILSPRGDTLMILKA 3359  
QY 344 ---T---F---Q--- 346  
Db 3360 QRPDGGLYTCVATNSYGDSEQDFKVVYVYKYPIDEITDQTPKAVAGGEIILKCPVLGNPT 3419  
QY 347 ---G---HT---N-E---VN---A-I-355 355  
Db 3420 PTVTKRGDDAVPNDRSRHTIWNVYDLKINSVTTEDAGQYSCIAVNEAGNLTHVAAEVIG 3479  
QY 356 ---K---W---D---P---T-360 360  
Db 3480 KPTFVRKGNLYEVIENDTITMCGVTSRPLPSISWFRGDKPVLYDYDRYSISPDGSHITI 3539  
QY 361 ---G---N---L---LA---S---C-S 368 368  
Db 3540 NKAKLSGGKYICRANEAGTSIDLILKILVPPKDKSNIIGNPLAIVARTIYLECPTS 3599  
QY 369 ---D---D-M--- 371  
Db 3600 GIPQPDVITKMGMDINWTDNRVILAQNNETFGIENVQVTDQGRYTCATNRGKASHDF 3659  
QY 372 ---TL---K---I---W---S---M-378 378  
Db 3660 SLDLVSPPEFIHGTOPIKREGTITLTCPILAEADIADQVMDVMSVTKDSRALDGLTD 3719

QY 379 ---K---Q---D--- 381  
Db 3720 NVDISDDGRKLTTISQASLENAGLYTCIALNRAGEASLEFKEVILSPVIDISRNDVQPOV 3779  
QY 382 ---N---C-V-H---L---Q-Q---H--- 390  
Db 3780 AVNQPTIMRCVAVTGHPFPSIKWLKNGKEVTDDEINRIVEQGVQLILRTDSDHAGKWSVCV 3839  
QY 391 ---N---KE---I-Y-TI---KW--- 399  
Db 3840 AENDAGVKELEMLVDVFTPEVWSKDNPIKALGETITLFCNAGSNPYQLKWKAGGSLI 3899  
QY 400 ---SP---T-G---P---GT---N--- 407  
Db 3900 FDSPDGARISLKGARLDIPLHKKTDVGDYTCQALNAAGTSEASVSDVLVPPINRDGID 3959  
QY 408 ---N---P--- 409  
Db 3960 MSPRLPAQOSLTLQCLAQKFPQMRWTLNGTALTHTSTPGITVASDSTFIQINNVSLSDK 4019  
QY 410 ---N-A-NLM---LA---S--- 417  
Db 4020 GVYTCYAEVAGSDNLMYVNVVQAPVINSNGTKQVIEGELAVIECLVEGYPAPQVSWLR 4079  
QY 418 ---A-SFDS--- 422  
Db 4080 NGRVETGVQVRYVTDGRMLTIIEARSLDSGIYLCSATNEAGSAQAQAYTLEVLVSPKII 4139  
QY 423 ---T---VR-L---W---DV--- 429  
Db 4140 TSPGVLTPSSGSKFSLPCAARGYPDPPIISWTLNGNDIKDGENGTIGADGTLHIEKABE 4199  
QY 430 ---D---R---G---I-C-I--- 435  
Db 4200 RHLIYECTAKNDAGADTLEFPQTVIAPKISTGNRYINGSEGTETVIKEIESESEFS 4259  
QY 436 ---H---T-L---T---K 440  
Db 4260 WSKNGVPLPSNNLISEDYKLIKILSTRLSDOGEYSCTAANKAGNATQKTNLNVGAPK 4319  
QY 441 ---H-Q---E---P-V-Y---SVAF--- 450  
Db 4320 IMERPRTQVVKHQDQVTLWCEASGVQPAITWYKDNELLTNTGVDETATTTKKKSVIFSSI 4379  
QY 451 SP---D---G--- 454  
Db 4380 SPQAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMNVSTNPNROTFLVSCNATGIPEP 4439  
QY 455 ---R---Y---LA---SGSF---D---K 464  
Db 4440 VISWMEDSNIAIONNEKYQILGTTLAINVLPPDDG-FYHCIAKSDAGQKIATRKLIYNK 4498  
QY 465 ---C---VHI---W---N--- 470  
Db 4499 PSRDPAPIWVECDKPKKTEYIMIDRGDTPDDNPOLLPKWDESSLNGSIAYRCMPGP 4558  
QY 471 ---T---Q---T-CA---LV--- 477  
Db 4559 RSRVTLLHAAPQFIVKPKNTAAIGAIVELRCSAAGPPHPTITWAKDGKLIEDSKFEIA 4618  
QY 478 ---H---S--- 479  
Db 4619 YSHLKVTLNSTSDGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKSSNQKNVAVITCYE 4678  
QY 480 ---Y-RG-T---G---GI--- 486  
Db 4679 RNQVSRGLTWYENGVPMPKNLAGIHFMMNGSLVILDTSSLKEGDLLEYTCKVNRNRHS 4738  
QY 487 ---FE---V-C---W-N---A--- 493  
Db 4739 IPHLTSAFEGVEVKTIDKVEVNGSDVLDCEVTDPLTTHVVTKNDQKMLDDDAIYV 4798  
QY 494 ---A---G---D---KV-G---AS--- 501

Db 4799 LPNLSVLLNVEKYDEGVKCVASNSIGKAFDDTQLNVYGGSSRRREAYKKNEDASTTTI 4858  
QY 502 -----A-----S-----DG----- 505  
Db 4859 TTTSPPTTTTETPLTTIIPALITLPAKQYPTDDYHEGSANDDGFPTTQDSLFEFNPPL 4918  
QY 506 ---SV---CV-----LD--- 511  
Db 4919 HPEISVVNTDCAGTINENGDCVDKDKTKHNLKLTGHNHCPEGFAMNPHTRICEDLDECA 4978  
QY 512 -----L-----R-----K 514  
Db 4979 FYQPCDFECINYDGGFCNCPGLGYELAEBCRDVNECESVRCBDGX 5024

RESULT 12  
US-10-369-493-6859  
; Sequence 6859, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6859  
; LENGTH: 5175  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-6859

Query Match 71.0%; Score 2596.6; DB 14; Length 5175;  
Best Local Similarity 9.0%; Pred. No. 2.1e-29;  
Matches 454; Conservative 48; Mismatches 10; Indels 4514; Gaps 390;

QY 1 M-----SI-----SS-----D-----E----- 7  
Db 1 MGRSPSLYGVLLGALLLATTCSVNDKNDPTGKSLAFVFDITGSMFDDLVQVREGAAK 60  
QY 8 ---V---N-----F---L---VY----- 13  
Db 61 IPKTVMAQREKLIYIMVPHDPYLGEIINTDSTYFMRQLSKVYVHGGDCPEKTLTG 120  
QY 14 -----R---Y-----L---QE---SG---F-----SHSAP-T 27  
Db 121 ILKALQISLPSPFIYVFTDARSKDHYLEDEVLNTIOEKQSSVVFVMTGDCGNRTHPGFRT 180  
QY 28 -----FG-I---KS-----H-ISO-----S-NI----- 39  
Db 181 YEKIAAASFQGFHLEKSDVSTVLEYVRHAVKQKVHLMYAEARERGGTVSRNIPVDKHL 240  
QY 40 -----N-----GA-----LV-----P-P----- 46  
Db 241 ELTISLGGKDDSDNLDLVLROPEGRVTKRLYSKEGGTIDLKNNVKILRLKDPSPGVWTV 300  
QY 47 -----A-----A-----LI---S---I----- 52  
Db 301 NTNSRLKHTIRVFGHGAVDKYGFASRLDRIELARPRVLNQDTYLLINMTGLIPPGTV 360  
QY 53 --I-----Q-----KGL-----Q-Y-----VE---AE- 63  
Db 361 GEIDLVDYHGHSYKAVASPHRTNPNWYFAGVPVPPKGLFFVRVQGYDEDNFEFMRAPT 420  
QY 64 ---V---S-I-----N-----ED---GTLF--- 73

Db 421 AIGSVIVGGPRAFMSPHIOEFVGRDLNLSTCVESASAYTIYVVKTGEDIIGGFLFYHNTD 480  
QY 74 -----D-----G-----R---P-I-----E----- 79  
Db 481 TSVWTIPELSLKADAGEYECRVISNNGNYSVKTRVETRESPPEIFGVRNVSVPLGEAAFLH 540  
QY 80 ---S---L-----SL-I-----DA----- 86  
Db 541 CSTRSAGEVEIRWTRYGATVFNPNPTNPTNCTGLKIHVTRADAGVYECMARNAGMST 600  
QY 87 ---VM-----P-DV----- 92  
Db 601 RKMRLDIMEPPSVKVTPOVVFYNNMREGVNLSCAMGDGPKPEVHWYFKRHLNDYKYQVG 660  
QY 93 Q-----T---R---Q---QAYRD---KLA-----Q----- 105  
Db 661 QDSKFLYIRDATHDEGTYECRAMSQQA-RDITDMLATPPKVEIIONKMVGRGDRV 719  
QY 106 -----Q-----Q-----A---A---A-----A- 111  
Db 720 SPECKTIRGKPHKIRWPKNGKDLIKPDDYIKINEQQLHMGAKDEKADAGAYSCVGENMAG 779  
QY 112 ---A-----A-----A-----A-----A----- 114  
Db 780 KDQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNNP 839  
QY 115 ---A-----A-----A-----ASQ----- 120  
Db 840 RYQLADGNLLITDAQIEDOGFTCIARTYQQSQSSTLMTVGLVSPVLGHVPPBEQLI 899  
QY 121 ---Q----- 123  
Db 900 EGQDLTLSCVVLGTPKPSIVMIKODKPVEEGPTIKIEGGSSLLRLRGGNPKDEGKYTCI 959  
QY 124 ---A-----A-----K-----N-----GE--- 128  
Db 960 AVSPAGNSTLHINVLKPKPEFYKPEGGIVFKPTISGMDEKHVAVVNSTHVDLDEGFA 1019  
QY 129 -----A-----N----- 133  
Db 1020 IPCVWSGTPTTITWYLDGRPTIPNSRDFTVTADNTLIVRKADKSYGVYTCOATNSAGD 1079  
QY 134 -E----- 136  
Db 1080 NEQKTTIRIMTPTMISPGOSSFNWVDDLTIPCDVYGDPKPVITWLLDDKPFTEGVVNE 1139  
QY 137 -G-----AH-----TI-A-N-----N----- 144  
Db 1140 DGSLLTIPNVNEAHRGFTTCHQAAGNDTRVTLTVHTTPTINAENQEKIALQNDIVLE 1199  
QY 145 -----HT---D-----MME-----V-----D- 152  
Db 1200 CPAKALPPVRLWYVEGEKIDSQLPHITIREGALVQLNVKLENTGVFCQVSNLAGEDS 1259  
QY 153 -----G-DV-----EIP---P-----NK-AV----- 163  
Db 1260 LSYTLTVHEKPKIISSEVPGVVDVVKGFTTIEICRATGVEVIRTWNKNGIDLKMDKKFS 1319  
QY 164 V---LR-----G-----H-----E-----S----- 170  
Db 1320 VDNLGLTRIYEAADKNDIGNYCNVVTNEAGTSQMTTHVDVQEPPIILPSTQNTNNAVVGDR 1379  
QY 171 -----E-----V---F-----I---C----- 175  
Db 1380 VELKCYVEASPPASVTWFRRGIAIGTDTKGYVVESDGLTVIOSASVEDATITYCKASNPA 1439  
QY 176 -----A----- 176  
Db 1440 GKAEANLQVTVIASPIKDIPDVVTOESIKESHPPSLYCPVFNPLPQISWYLNKPLIDD 1499  
QY 177 ---W-----N----- 178

Db 1500 KTSWKTSDDKRKLHVFKAKITDGSYKVCARNAAGEGSKSQVEVIVPLNLDESKEYKKV 1559  
Qy 179 -----PVS----- 181  
Db 1560 FAKEBEVTLGCPVSGFPVPOINWVDTGVVEPGKYKGATLSNDGLTLHFDSVSVKQEG 1619  
Qy 182 -----D-L-LA-----S 186  
Db 1620 NYHCAQSKGNILIDVELSVLAVPIVGEDDNLVFLGKISLSDQLQTESDDKTTFW 1679  
Qy 187 -GS-----G-----DS-----T-----A-R----- 194  
Db 1680 INGSERDPNVQIPSDGHRLYITDAKPENNGKVMCRVNTSAGKAERTLTLVLEPPVF 1739  
Qy 195 -----IW-----N-----LS----- 199  
Db 1740 EPVPEANOKLIGNPIILOQCVTGNPKPTVIWKIDGNDVDKMLFDBESLSLLRIEKLTK 1799  
Qy 200 -----EN-----S-----T-----S-GS-----T- 207  
Db 1800 SAQISCTAENKAGTASRDFFIQNLAAPTFKNEGDOETIPRESEITILDCPVSLGDFQITW 1859  
Qy 208 -Q-----L-VL-----R-H-----C-----IR 216  
Db 1860 MKQGLPLTENDAIFTLNTRTLILNANRDHEDIYTCVANNTAGQVSKDFDVWVQVLPKIK 1919  
Qy 217 -----EG-----G-----Q-DVP----- 223  
Db 1920 NAVVLEINGBEILTCDAENPTPTAKWDFNOGLPKCAVFNHNNHTVVANNVTKYHT 1979  
Qy 224 -----SNK-----DV-----T-S-L-----D----- 232  
Db 1980 GVKCYATNKVGQAVKTIHVHVRKPRFESGLTESELTVNLTRSLITLCECDVDAIGVIS 2039  
Qy 233 W-N-----SE-G-TL-----L-----A-T-----GSY----- 245  
Db 2040 WTVNGKFLAETDGVQTLAGRFLHIVSAKTDHGSYACTVTNEAGVATKTNFLVQVPP 2099  
Qy 246 -----DG-----F----- 248  
Db 2100 TIVNEGYYTVIENSLVLPCEVTGKPNPVVTKGRPVGDLSKVQVLSGQOQKIVHA 2159  
Qy 249 -----A-----R----- 250  
Db 2160 EIAHKGSYICWAKNDVGTAEISFDVDIITRPMIQKGIKNIVTAIKGALPKFCPIDDDKN 2219  
Qy 251 -----IW-----T-----K-D-G-N----- 257  
Db 2220 FKQIWLNRNYQPIDLEADARITRLSNDRRLTILNVTENDEGQYSCRKNDAGENSFDF 2279  
Qy 258 -----L-----A-----S-TL-----QO-H----- 265  
Db 2280 KATVLPPTIIMLDKKNKTAVEHSTVTLSCPATGKPEPDITWFKOGEAIIHENIADIIP 2339  
Qy 266 -----K-G-P-----I----- 269  
Db 2340 NGELNGQLKTRIKEGDAKYTCADNSAGSVEQDVNVNVTIPKIEKDGIPSDYESQ 2399  
Qy 266 -----K-G-P-----I----- 269  
Db 2400 NERVWISCPVYARPAKITWLGAKPLQSDKFKVTSANGOKLYLFKLRETDSSKYTCIAT 2459  
Qy 270 -----F-----A-----L-----K-----W-----N----- 275  
Db 2460 NEAGTKRDFKVMVLVAPSEDFENIVRITVNSGNSTLHCPAKGSPFITWLGKNAI 2519  
Qy 276 -----K----- 276  
Db 2520 EPNDRVVFFDAGRLQISKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPPIDGERREA 2579  
Qy 277 -----K-G-----N-----FI-----LS----- 283  
Db 2580 VAVIEGFSSELCDSNSTGVDVWQKDLGTINQDTRLGRGFSFIOIPSGKMKMSLSARKSD 2639

Qy 284 -----AG-----V-----DK-T-----TI-----I----- 292  
Db 2640 SGRYTCIVRNPAEAKRLDFVANDPPSISDELSSANIQTIVPPYVPEINCVVSGSPHPK 2699  
Qy 293 -----W-----D-A-----H-----T-----G----- 298  
Db 2700 VYWLFDKPLPDSAAVELTNNGETLKIVRSQVEHAGTYTCEAQNNVVGKARKDFLVRVTA 2759  
Qy 299 -----E-----AK-----Q----- 302  
Db 2760 PPHFEKERBEVVARVGDITMLLTCAESSVPLSSVYVWHAHDESQVNGVITSKYAANEKTLN 2819  
Qy 303 -----Q-----P----- 304  
Db 2820 VTNIQLDDGFFYYCTAVNEAGITKFKFLIVITPYFLDQOKLYPIILKRLTLDCSATG 2879  
Qy 305 -----P-----F----- 306  
Db 2880 TPTPTILFMKDGKRLNESDEVDIIGSTLVIDNPQKEVEGRTYCIAENKAGRSEKDMVVEV 2939  
Qy 307 -----H----- 308  
Db 2940 LLPPKLSKEWINVEVOAGDPLTLECIPTEDTSGVHITWSRQFGKQGLDMRAQSSDKSL 2999  
Qy 309 -----A-P-A-L-----D-----V-D----- 315  
Db 3000 YIMQATPEDADSYSCIAVNDAGAEAVFQVTVNTPPKIFGDSFSTTEIVADTTLEIPCRT 3059  
Qy 316 -----W----- 317  
Db 3060 EGIPPEISWFLDGKPILEMPGVYTKQDLSLRIDNIKNQEGRYTCVAENKAGRAEQDT 3119  
Qy 318 -----S-----N-----N-----T-----P----- 322  
Db 3120 YVEISEPPRVWASEVMRVVEGRQTTIRCEVFGNPEPVVWNLKDGSPYTSLLQFSTKLS 3179  
Qy 323 -----A-----SC-----S-----TD-----M-C- 330  
Db 3180 YLHLRETTIADGGTYTCIATNKAGESQTTDDEVLPVPRIEDBEERVLOKSEGNTYVWHQ 3239  
Qy 331 -----I-----HV-----C-----K 335  
Db 3240 VTGRPVYVTKWNGKEIEIQFNPLHIRNATRADEGKYSIASNEAGTAVADFLIDVFTK 3299  
Qy 336 -----L-GO-----D-----RP-----I-K- 343  
Db 3300 PTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRPFNMNDNIIILSPRGDTMLKA 3359  
Qy 344 -----T-----F-----Q----- 346  
Db 3360 QRFDGGLYTCVATNSYGDSEQDFKVVNTKPYIDETIDOTPKAVAGEIILKCPVLGNPT 3419  
Qy 347 -----G-----HT-----N-----E-----VN-----A-I- 355  
Db 3420 PTVTKRGDDAVPNDNRHTIIVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVIG 3479  
Qy 356 -----K-----W-----D-----P-----T- 360  
Db 3480 KPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYLDYRYSIGPDGSHITI 3539  
Qy 361 -----G-----N-----L-----LA-----S-----C-S 368  
Db 3540 NKAKLSDGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPIS 3599  
Qy 369 -----D-----D-M----- 371  
Db 3600 GIQPDVITWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGKGASHDF 3659  
Qy 372 -----TL-----K-----I-----W-----S-----M- 378  
Db 3660 SLDLVSPPEFDIHGTOPTIKREGDTITLTCPIKLAEDTADQVMDVSWTKDSRALDGLD 3719

QY 379 -----K-Q-----D----- 381  
Db 3720 NVDISDDGRKLTTISQASLENAGLYCTIALNRAGEASLEFKVELTSPVIDISNDVQPOV 3779  
QY 382 --N--C-V-H-----D--L--Q-Q--H----- 390  
Db 3780 AVNQPTIMRCVGTGHPFPIKMLKNGKEVTDENIRIVGQVQLRLRTSDSHAGKWCV 3839  
QY 391 --N--KE-----I-Y-TI-----KW----- 399  
Db 3840 AENDAGVKELEWLDVFTPPVSVKSDNPICALGETITLFCNAGSNPYQLKWAKGSLI 3899  
QY 400 --SP--T-G--P-----GT-----N----- 407  
Db 3900 FDSPPGARISLKGARLDIPLHKKTVDGDTYCOALNAAGTSEASVSDVLVPPPEINRDGID 3959  
QY 408 -----N-----P----- 409  
Db 3960 MSPLPAQOSLTLQAOQKPPQMRWTLNGTALTHTSTPGITVASDSTFIQINNVSLSDK 4019  
QY 410 -----N-A--NLM-----LA-----S-- 417  
Db 4020 GUYTCYAEENVAGSDNLMYNDVQAPVINSNGTKQVIEGELAVIECLVEGYPAQVSWLR 4079  
QY 418 -----A-SFDS----- 422  
Db 4080 NGRVETGVQVRYVTDGRMLTIEARSLDSGIYLCSATNEAGSAQAQVTLVLVSPKII 4139  
QY 423 -----T-----VR-L--W-----DV----- 429  
Db 4140 TSTPGVLTSPSGSKFSLPCAVRGPDPIISWTLNGNDIKDGENGTIGADGTLHIEABE 4199  
QY 430 -----D-----R-----G-----I-C-I----- 435  
Db 4200 RHLIYECTAKNDAGADTLFPVQTVIAPKISTGNGRYINGSEGTETVICKIESESSEFS 4259  
QY 436 -----H-----T-L-----T-----K 440  
Db 4260 WSKNGVPLPLSNLIFSDYKLIKILSLRLSDQGEYSCTAANKAGNATQKTNLNVGAPK 4319  
QY 441 -----H-Q-----E-----P-V-Y-----SVAP-- 450  
Db 4320 IMERPRTQVHKGDQVTLWCEASGVQPAPITWYKONELLTNTGVDSTATTKKSVIFSSI 4379  
QY 451 SP-----D-----G-----G----- 454  
Db 4380 SPSQAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMVSTNPROTVFLSCNATGIPRP 4439  
QY 455 -----R-----Y-----LA-----SGSF-----D-----K 464  
Db 4440 VISWMRDSNIAIQNNEKYQILGTTLAIRNVLPDDG-FYHCIAKSDAGQKIATRKLIYNK 4498  
QY 465 -----C-----VHI-----W-----N----- 470  
Db 4499 PSDRPAPIWVECKEKKPKTEYIMIDRGDTPDDNPQLLPKWQVEDSSLSNGSIAYRCMPGP 4558  
QY 471 -----T-----Q-----T-----CA-----LV----- 477  
Db 4559 RSRVTLLHAAPQFIVKPKNTTAAIGAIVELRCSAAGPPHPTITWAKOGKLIEDSKFEIA 4618  
QY 478 --H-----S----- 479  
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSAPINVNDNILPTPKSSNQKNVAVITCYE 4678  
QY 480 -----Y-RG-T--G-----GI----- 486  
Db 4679 RNQAYSRLGTWYGVMPKPNLAGIFHMNNGSLIVLDTSSLKEGOLEYLTCVKNRRRHS 4738  
QY 487 -----FE-----V-C-----W-N-----A--- 493  
Db 4739 IPHLTSAFEGVPEVKTIDKVEVNGSDSVLDCVETSDPLTTHVWTKNDQKMLDDAIYV 4798  
QY 494 -----A--G--D--KV--G-----AS--- 501

Db 4799 LPNNSIVLLNVEKYDEGVYKCVASNSIGKAFDDTQNLNVYGGSSRRREAYKKENEDASTTTI 4858  
QY 502 -----A-----S--DG----- 505  
Db 4859 TTTSTPTTTTETPLTTIIPALITILPAKOYPTDDVYHEGSANDDGFPTTQDSLFEFNPPL 4918  
QY 506 --SV--CV-----LD--- 511  
Db 4919 HPEISVWNTDCAGTINENGDCVDKQKTHNLKILLTGENHCPEGFAMNPHPTICEDLDECA 4978  
QY 512 -----L-----R---K 514  
Db 4979 FYQPCDFECINVDGFGQCNCPGLGYELAEGRCDVNECESVRCEGDK 5024

RESULT 13  
US-10-369-493-6861  
; Sequence 6861, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6861  
; LENGTH: 5175  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-6861

Query Match 71.0%; Score 2596.6; DB 14; Length 5175;  
Best Local Similarity 9.0%; Pred. No. 2.1e-29;  
Matches 454; Conservative 48; Mismatches 10; Indels 4514; Gaps 390;  
QY 1 M-----SI-----SS-----D-----E--- 7  
Db 1 MGRSPSWLYGVLGLLLATTCSVNDKNDPTGKSSLAFFDITGSMFDDLVQVREGAAK 60  
QY 8 ----V-----N-----F--L--VY----- 13  
Db 61 IFKTVAQREKLIYNYVPHDPYLGELIINTDSTYFMRQLSKYVYVHGGDCPEKTLTG 120  
QY 14 -----R--Y-----L--OE--SG--F-----SHSAP-T 27  
Db 121 ILKALQISLPSFIYVFTDARSKDYHLEDEVLTITQEQSSVVFVMTGCGNRTHPGFRT 180  
QY 28 -----FG-I---KS-----H-ISO-----S-NI----- 39  
Db 181 YEKIAAASFGQVHFLEKSDSVTVLEVRHAVKQKVHLMYEARERGTVSRNIPVDKHLS 240  
QY 40 -----N-----GA-----LV---P-P--- 46  
Db 241 ELTISLGDKDDNLDIVLRDPGRTVDKRLYSKEGGTIDLKNVKLIRLKDPSFGVWTV 300  
QY 47 -----A-----A-----LI--S--I----- 52  
Db 301 NTNSRLKHTIRVFGHGAVDKYGFPASPLDRIELARPRVLNQDTYLLINMTGLIPPGTV 360  
QY 53 --I-----Q-----KGL---Q-Y---VE---AE- 63  
Db 361 GBIDLVDYHGHSLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGVDBDNFYEFRIAPT 420  
QY 64 -----V-----S-I-----N-----ED---GTLF----- 73

Db	421	AIGSVIUGFRAPMSPIHOEFVGRDNLNLSCTVESASAYTIYVWKTGEDIIIGPLFHNHTD	480	Db	1500	KTSWKTSDDKRLHFKAKITDSGVYKCVARNAAGEGSKSQFVEVIVPLNLDSEKYYKKV	1559
Qy	74	-----D-----G-----R-----P-I-----E-----	79	Qy	179	-----PVS-----	181
Db	481	TSWVTIPELSLKDAGEVECRVSNNGYSVKTRETPESPPEIFGVRNVSVPLGEAFLH	540	Db	1560	FAKEGEVTLGCPVSGFPVPOINWVDTGVVPEPKYKAGATLSNDGLTLHFDSVSVKQEG	1619
Qy	80	-----S-----L-----SL-I-----DA-----	86	Qy	182	-----D-L-LA-----	186
Db	541	CSTRSAGEVEIRTRYGATVFNPNTERNPTNGTLKHVHTRADAGVYECMARNAGMST	600	Db	1620	NYHCVAQSKGNILDDVELSVLAVPIVGEDDNLVFLGKDISLSCDLQTESDDKTFFVMS	1679
Qy	87	-----VM-----P-DV-----V-----	92	Qy	187	-----GS-----G-----DS-----T-----A-R-----	194
Db	601	RKRLDIMEPPSVKVTQDQVYFNMRGVNLSCEAMGDPKPEVHWYFKGRLHLLNDYKQVG	660	Db	1680	INGSSEDRPDNVQIPSDGHRLYITDAKPENNKYMCRVTSAGKABRTLTLDVLEPPVFW	1739
Qy	93	Q-----T-----R-----Q-----QAYRD-----KLA-----Q-----	105	Qy	195	-----EN-----S-----T-----S-GS-----T-----	199
Db	661	QDSKELYIRDATHDEGTVECRAMSQAQA-RDITDLMLATPPKVEIIQNMWVGRGDRV	719	Db	1740	EPVFEANOKLIGNPIILOCQVTGNPKPTVIWKIDGNDVDKSWLFDSELSLLRIEKLTK	1799
Qy	106	-----Q-----Q-----Q-----A-----A-----A-----	111	Qy	200	-----SAQISCTAENKAGTASRDFFIQAAPTFKNEGDOETIFRESEITITLDCPVSLGDFQITW	1859
Db	720	SPECKTIRGPKPKIRWFKNGKDLIRKPDYIKINEGOLHIMGAKDEBAGAYSCVGNMAG	779	Qy	208	-----Q-----L-VL-----R-H-----C-----	216
Qy	112	-----A-----A-----A-----A-----A-----	114	Db	1860	MKQGLPLTENDAIFTLDNTRLTILNARDHEDIYTCVANNTAGQVSKDFVVDVVLPLKIK	1919
Db	780	KOVQVANLSVGRVPTTIESPHTVRNIERQVTLQCLAVGIPPEIEWKGNVLLATLNNP	839	Qy	217	-----EG-----G-----Q-DVP-----	223
Qy	115	-----A-----A-----A-----ASQ-----	120	Db	1920	NAVVTLEINEGEEIILTCDAEGNPTTAKWDPNQDLPKEAVFVNNHTVVVNNVTKYHT	1979
Db	840	RYTQADGNLLITDAQIEDQOQTCIARNTYGOQSOSTTLMVTLGVLSPVLGHVPPEQLI	899	Qy	224	-----SNK-----DV-----T-S-L-----D-----	232
Qy	121	-----Q-----Q-----GS-----	123	Db	1980	GVYKCVATNKVQAVKTIINVHVRTPKPFESGLTESELTVNLTSTRITLEDVDVDAIGVGIS	2039
Db	900	EGQDLTLCVVVLGTPKPSIWIWKDKPVEBEGPTIKIEGGSLRLRGGNPKDEGKYTCI	959	Qy	233	W-N-----SE-G-TL-----L-----A-T-----GSY-----	245
Qy	124	-----A-----A-----K-----N-----GE-----	128	Db	2040	WTVNGKPLAETDGVQTLAGGRFLHIVSAKTDHGSACTVTNEAGVATKTNLFOVQPP	2099
Db	960	AVSPAGNSTLHINVQLIKPEFVYKPEGGIVFKPTISGMDEKHVAVVNSTHVDLDEGFA	1019	Qy	246	-----DG-----F-----	248
Qy	129	-----NT-----A-N-G-----	133	Db	2100	TIVNEGVEYTIENNSLVLPCEVTGKPNPVVTWKDGRPVGLKSVQVLSGQFKI VHA	2159
Db	1020	IPCVSGTPPPIITWYLDGRPTPNSRDFTVTADNTLIVRKADKSYGVVTCQATNSAGD	1079	Qy	249	-----A-----R-----	250
Qy	134	E-----E-----E-N-----	136	Db	2160	ETAHKGSYICMAKNDVGTAEISFDVDIITRPMIOGKIKNIVTAIRKGGALPFKCPIDDDKN	2219
Db	1080	NEQKTIIRIMTPMISPGQSSFNWVDDLFTIPCDVYGDPKPVITWLLDDKPTGEGVNE	1139	Qy	251	-----IW-----T-----K-D-G-N-----	257
Qy	137	G-----AH-----TI-A-N-----N-----	144	Db	2220	FKQIILWLRNYQPIDLEADARITLNSDRRLTILNVTEDEGOYSCRKNDAGENSFDF	2279
Db	1140	DGSLTIPNVNEAHRGTTCHAQNAAGNDTRVTTLTVHTTPTINAENQEKIALQNDIVLE	1199	Qy	258	-----L-----A-----S-TL-----GO-H-----	265
Qy	145	-----HT-----D-----MME-----V-----D-----	152	Db	2280	KATLVPPPTIIMLDKDKNKTAVEHSVTLSCTPATGKPEPDITWFKDGEAHIENIADIIP	2339
Db	1200	CPAKALPPPRLWTEGEKIDSQLPHITREDGALVQLQNVKLENTGTVFCQVSNLAGEDS	1259	Qy	266	-----K-----	265
Qy	153	-----G-DV-----EIP-----P-----NK-AV-----	163	Db	2340	NGELNGNOLKITRIKEGDAGKYTCCEADNAGSVGEQDWNVNVTIPKIEKDGPISDYESQ	2399
Db	1260	LSYTLTWHEKPIIIEVPGVVDVVKGFTIIEPCRATGVPEVIRTNKNGIDLWKDEKPS	1319	Qy	266	-----K-G-P-----I-----	269
Qy	164	V-----LR-----G-----H-----E-----S-----	170	Db	2400	NERVISCVPYARPPAKITWLKAGPLQSDKFKVTSANGOKLYLFKLRETDSSKYTCIAT	2459
Db	1320	VDNLGTLRIEADKNDIGNYCVVNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVGDR	1379	Qy	270	-----F-----A-----L-----K-----W-----N-----	275
Qy	171	-----E-----V-F-----I-C-----	175	Db	2460	NEAGTKDRDFKVSMLVAPSFDEBNIVRRITVNSGNPSTLHCPAKGSPSTITWLKDGNAI	2519
Db	1380	VELKCYEASPPASVTWFRREGIAIGTDTKGYVWESDGLTVIQSASVEDATIYTCKASNP	1439	Qy	276	-----K-----	276
Qy	176	-----A-----	176	Db	2520	EPNDRYVFFDAGRQLQISKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPVVDGERREA	2579
Db	1440	GKAENLOVTVIASPDIKDPDVVTQESIKESHPSLYCPVPSNPLPOISWVNDKPLIDD	1499	Qy	277	-----K-G-N-----F-----LS-----	283
Qy	177	-----W-----N-----	178	Db	2580	VAVIEGFSSELFCDSNSTGVVDVWQDGLTINQDITLGRGDSFIQIPSSGKMSFLSARKSD	2639



284 QY ---AG---V---DK-T---TI---I--- 292  
2640 DB SGRYTCIVRPAGEARKLFDFAVNDPPSISDELSSANIQTIVPPYVEINCVVSGSPHPK 2699  
293 QY ---W---D-A---H---T---G--- 298  
2700 DB VYWLFDKLEPDASAAYELTNNGETLKIVRSQVEHAGTYTCEAQNNGVKARKDFLVRVTA 2759  
299 QY ---E---AK---Q--- 302  
2760 DB PHFEKERBEVARVGDTHLLTCNAESSVPLSSVYVYHHADESQVQVITTSKYAANEKTLN 2819  
303 QY ---Q---F--- 304  
2820 DB VTNIQLDDSGFYCTAVNEAGITKFKLIVETPYFLDQQLYPILGKRLTLDSCATG 2879  
305 QY ---P---F--- 306  
2880 DB TPPPTILFMKDGKRLNESDEVDIIGTLVIDNPQKEVEGRYTCIAENKAGRSEKMMMEV 2939  
307 QY ---H---S--- 308  
2940 DB LPPPKSKWINVEVOAGDPLTLECPIDTSGVHITWSRQFGKQGLDMRAQSSDKSL 2999  
309 QY ---A-P-A-L---D---V-D--- 315  
3000 DB YIMQATPEDADSVCIAVNDAGAEAVFQVYVNTTPPKIFGDSFSTTEIVADTTLBIPCRT 3059  
316 QY ---W---Q--- 317  
3060 DB EGIPPEISWFLDKGPILEMPGVYTKQGLSLRIDNIKENQEGRYTCVAENKAGRAEQDT 3119  
318 QY ---S---N---T---F--- 322  
3120 DB YVEISPPRVWASEVMRVVEGRQTTIRCEVFGNPEPVVNLKDGEPYTDLLQFSTKLS 3179  
323 QY ---A---SC---S---TD---M-C- 330  
3180 DB YLHLRETTLDGGTYTCIATNKAGESQTTDVEVLVPPRIEDBEERVLOQKEGNTYVHVCQ 3239  
331 QY ---I---HV---C---K 335  
3240 DB VTGRPVYVTKRNGKEIEQFNPVLHNRATRADEKGYSCIASNEAGTAVADFIDVFTK 3299  
336 QY ---L-GQ---D---RP---I-K- 343  
3300 DB PTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRPFNMNDIILSPRGDTLMILKA 3359  
344 QY ---T---F---Q--- 346  
3360 DB QRPDGLYTCVATNSYGDSEQDFKVVYTKPYIDETIDQTPKAVAGEIILKCPVLGNPT 3419  
347 QY ---G---HT---N---E---VN---A-I- 355  
3420 DB PTVTWKRGGDAVNDPSRRTIVANNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVIG 3479  
356 QY ---K---W---D---P---T- 360  
3480 DB KPTFVRKGNLYXEVIENDTITMDCGVTSRPLPSISWFRGDKPVYLYDRYSISPDGSHITI 3539  
361 QY ---G---N---L---LA---S---C-S 368  
3540 DB NKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIYLECPIIS 3599  
369 QY ---D---M--- 371  
3600 DB GIPQPDVIWTKMGWDINWDSRILVAQNNETFGIENVQVTDQGRYTCATNRGKASHDF 3659  
372 QY ---TL---K---I---W---S---M--- 378  
3660 DB SLDVSPPEFDIHGTOPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGLDLD 3719

379 QY ---K-Q---D--- 381  
3720 DB NVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEPKVEILSPVIDISRNDVQPOV 3779  
382 QY ---N---C-V-H---D---L---Q-Q---H--- 390  
3780 DB AVNQPTIMRCAVTHGPPFPIKWLKNGKEVTDENIRIVEQGVQLILRTDSDHAGKMSCV 3839  
391 QY ---N---KE---I-Y-TI---KW--- 399  
3840 DB AENDAGVKELENVLDVFTPPVVSUKSDNPIKALGETITILFCNASGNPYQLKWAKGSLI 3899  
400 QY ---SP---T-G---P---GT---N--- 407  
3900 DB FDPDGCARISLKGARLDIPLHLKKTVDGYTCQALNAAGTSEASVSDVLVPPPEINRDGID 3959  
408 QY ---N---P--- 409  
3960 DB MSPRLPAQOSLTLQCLAQKVPQMRWTLNGTALTHTSTPGITVADSTFIQNNVLSDK 4019  
410 QY ---N-A-NLM---LA---S--- 417  
4020 DB GYTYCAENVAGSDNLNMYNDVVPAPVISNGGKTQVIEGELAVIECLVSGYPAPQVSWLR 4079  
418 QY ---A-SFDS--- 422  
4080 DB NGNRVETGVQVRYVTDGRMLTIEARSILDSGIYLCSATNEAGSAQAYTLEVLVSPKII 4139  
423 QY ---T---VR-L---W---DV--- 429  
4140 DB TSTPGVLTSPSSGSKFSLPCAARGPDPPIISWTLNGNDIKDNGHGTIGADGTLHIEABE 4199  
430 QY ---D---R---G---I-C-I--- 435  
4200 DB RHLIIECTAKNDAGADTLEFPVQTVAPKISTGSHRYNGSEGETVIEKIESESESEFS 4259  
436 QY ---H---T-L--- 440  
4260 DB WSKNGVPLLPNNLIFSEDYKLIKILSTRLSQGEYSCTAANKAGNATQKTNLWGVAPK 4319  
441 QY ---H-Q---E---P-V-Y---SVAF--- 450  
4320 DB IMERPRTQVHVKGQDQVTLWCEASGVFPQPAITWYKONELLTNTGVDETATTKKSVFSSI 4379  
451 QY ---SP---D---G--- 454  
4380 DB SPQAGVYTCKAENWVASTEEDIDLIVMIPPEVPERMVNTPROTFLVSCNATGIPEP 4439  
455 QY ---R---Y---LA---SGSF---D---K 464  
4440 DB VISWMRDSNIAIQNNEKYQILGTTLAIRNVLPDDG-FYHCIAKSDAGQKIATKRLIVNK 4498  
465 QY ---C---VHI---W---N--- 470  
4499 DB PSDRPAPIWVEDEKPKKTYMIDRGDTPDNPOLLPMKQVDESSLNGSIAYRCMPGP 4558  
471 QY ---T---Q---T---CA---LV--- 477  
4559 DB RSRVTLLHAAPQFIVKPKNTAAIIVELRCSAAGPPHPTITWAKDKLIEDSKFEETA 4618  
478 QY ---H---S--- 479  
4619 DB YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSAPINVDNNILPTPKSSNQKNVAVITCYE 4678  
480 QY ---Y-RG-T---G---GI--- 486  
4679 DB RNQAYSRLGTWEYNGVPMPKNLAGIHFMMNGSLIVLDTSSLKEGDLLEYTCVKVRNRHRS 4738  
487 QY ---FE---V-C---W-N---A--- 493  
4739 DB IPHLTSAFEGVEVKTIDKVEVNGSDVLDCEVSDPLTTHVVWTKNDQKMLDDDAIYV 4798  
494 QY ---A---G---D---KV---G---AS--- 501



Db 1200 CPAKALPPVRLTYEGEKIDSLIPHTIREDGALVLQNVKLENTGVFVQVSNLAGEDS 1259  
QY 153 -----G-DV-----EIP-----P-----NK-AV-----163  
Db 1260 LSYTLTVHEKPIIIEVPGVVDVVGFTIBIPCRATGVPDEVIRTNKNGIDLKMDEKKFS 1319  
QY 164 V-----LR-----G-----H-----E-----S-----170  
Db 1320 VDNLTGLRIYEADKNDIGNVNVVNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVGDR 1379  
QY 171 -----E-----V-----F-----I-----C-----175  
Db 1380 VELKCYEASPPASVTWFRGIAIGTDTKGYYVESDGLTVIQSASVEDATIYTCASNPA 1439  
QY 176 -----A-----176  
Db 1440 GKAEANLQVTVIASPIKDPDVVTOBSIKESHFPFSLYCPVFSNPLPQISWYLNDRPLIDD 1499  
QY 177 -----W-----N-----178  
Db 1500 KTSWKTSDDKRLHVFKAKITDSGVYKCVARNAAGEGSKSFQVEVIVPLNLDSEKYYKKV 1559  
QY 179 -----PVS-----181  
Db 1560 FAKEGEVTLGCPVSGFPVPPQINWVVDGTVEPGKYKATLSNDGLTLHFDSSVVKQEG 1619  
QY 182 -----D-----L-----LA-----S-----186  
Db 1620 NYHCVAQSGNILDIDVELSVLAVPIVGEDDNLVFLGKDLSLSDLOTESDDKTTFFVWS 1679  
QY 187 -----G-----DS-----T-----A-----R-----194  
Db 1680 INGESDRPDNVQIPSDGHLIYTDKAPENNGKYMCRVNTSAGKABRTITLDVLEPPVVF 1739  
QY 195 -----IW-----N-----LS-----199  
Db 1740 EPVFEANQKLGNNPILOQVGNPKPTVIWKIDGNDVDKSWLPDESLSLLRIEKLTKG 1799  
QY 200 -----EN-----S-----T-----S-----GS-----T-----207  
Db 1800 SAQISCTAENKAGTASRDPFIQIAAPTAKNEGDQETIPRESEITLDCPVSIGDFQITW 1859  
QY 208 -----Q-----L-----VL-----R-----H-----C-----IR-----216  
Db 1860 MKQGLPLTENDAIFLIDNTRLTILNARDHEDIYTCVANTAGQSKDFDVVVQVLPKIK 1919  
QY 217 -----EG-----G-----Q-----DVP-----223  
Db 1920 NAVVLEINEGEBIILTCDAEGNPTAKWDFNQGLPKEAVFVNNHTVWNNVTKYHT 1979  
QY 224 -----SNK-----DV-----T-----S-----L-----D-----232  
Db 1980 GUYKCVATNKGOAVKTIINVHRTKPRFESGLTESELTVNLTRISITLECDVDDAIGVGIS 2039  
QY 233 W-N-N-----SE-----G-----TL-----L-----A-----T-----GSY-----245  
Db 2040 WTVNGKPFLEATDGVQTLAGGRFLHIVSAKTDHGSYACTVINEAGVATKTNLFVQVPP 2099  
QY 246 -----DG-----F-----248  
Db 2100 TIVNEGGEVTVIENNSLVLPCVETGKPNFVVTWKDGRPVGDLKSVQVLSGQPKIVHA 2159  
QY 249 -----A-----R-----250  
Db 2160 ETIAHGSYICMAKNVGTAEISFDVDIITRPMIQKGIKNIVTAIKGALPFPKCPIDDDKN 2219  
QY 251 -----IW-----T-----K-----D-----G-----N-----257  
Db 2220 FKQIILWLRNYQPIDLEADARITLSNDRRLTILNVTEENDEGOYSCRVDNAGENSDF 2279  
QY 258 -----L-----A-----S-----TL-----GO-----H-----265

Db 2280 KATVLVPPTIIMLDKKNKTAVEHSTVTLSCPATGKPEPDIITWFKDGEAIHieniADIIP 2339  
QY 266 -----266  
Db 2340 NGELNGNQLKITRIKEGDAGKYTCADNSAGSVQEDVNNVNTIPKIKXDGPDSYESQ 2399  
QY 266 -----K-----G-----P-----I-----269  
Db 2400 NERVVISCVPYARPPAKITWLKAGKPLQSDKFVKTSANGQKLYLFKLRETDSSKYTCIAT 2459  
QY 270 -----F-----A-----L-----K-----W-----274  
Db 2460 NEAGTDKDRDFKVSMLVAPSFBNIVRRITVNSGNPSTLHCPAKGSPSPTITWLKDNAI 2519  
QY 275 -----N-----K-----276  
Db 2520 EPNDRYVFFDAGRQLOISCTEGSDQGRYTCIATNSVGSDDLENTLEVIIPPVIDGRREA 2579  
QY 277 -----K-----G-----N-----FI-----LSA-----284  
Db 2580 VAVIEGFSSELCDSNSTGVDVEMQKDLTINODTLRGDSFIOIPSSGKKMSFLSARKSD 2639  
QY 285 -----G-----V-----DK-----T-----TI-----I-----292  
Db 2640 SGRYTCIVRNPAAGEARKLDFAVNDPPSISDBLSANIQTIVPYYPVEINCVVSGSPHK 2699  
QY 293 -----W-----D-----A-----H-----T-----G-----298  
Db 2700 VYWLFDKPLEPDSAAVELTNNGETLKIVRSQVEHAGTYTCEAQNNVGKARKDPLVRVTA 2759  
QY 299 -----E-----AK-----Q-----302  
Db 2760 PPHFEKEREEVAVGDTMLLTCAESSVPLSSVYVHHADESQNGVITSKYAANEKTLN 2819  
QY 303 -----Q-----F-----304  
Db 2820 VTNIQLDDEGFYCTAVNEAGITKFKFLIVETPVFLDQKLYPIILGKRLTLDGSATG 2879  
QY 305 -----P-----F-----306  
Db 2880 TPTPTILFMKQKRLNESDEVDIIGTLVIDNPQKEVEGRYTCIAENKAGRSEKMMVEV 2939  
QY 307 -----H-----S-----308  
Db 2940 LLPPKLSKEWINVEVQAGDPLTLECPIDTSGVHLTWSQFGKQDGLDMRAQSSSKSL 2999  
QY 309 -----A-----P-----A-----L-----D-----V-----D-----315  
Db 3000 YIMQATPEDADSYSCIAVNDAGBAVFOVTNTPPKIFGDSFSTTEIVADTTLTIPCRT 3059  
QY 316 -----W-----Q-----317  
Db 3060 EGIPPEISWFLDGPILFEMPGVTKYQGLSLRINIKPNQEGRYTCVAENKAGRAEQDT 3119  
QY 318 -----S-----N-----N-----T-----F-----322  
Db 3120 YVEISEPPRVWASEVMRVVEGRQTIRCEVFNPEPVVNLKDGEPYTDLLQFSTKLS 3179  
QY 323 -----A-----SC-----S-----TD-----M-----C-----330  
Db 3180 YLHLRETTILADGGTYTCIATNKAGESQTTTDEVLVPPRIEERVLQKEGNTYMHQC 3239  
QY 331 -----I-----HV-----C-----K-----335  
Db 3240 VTGRPVVYVTKRNKBEQFNPVLHIRNATRADEGKYSICIASNEAGTAVADFLIDVFTK 3299  
QY 336 -----L-----GQ-----D-----RP-----I-----K-----343  
Db 3300 PTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGGPFNMDNILLSPRGDTMLKA 3359  
QY 344 -----T-----F-----Q-----346  
Db 3360 QRFQDGLYTCVATNSYGDSEQDFKVNVTYPYIDETIDQTPKAVAGGEIILKCPVLGNPT 3419

QY 347 ---G---HT---N---E---VN---A---I--- 355  
Db 3420 PTVTKRGDDAVPNDSDRHTIYNNVDLKINSVTTEDAGQYSCIAYNEAGNLTHYAAEVIG 3479  
QY 356 ---K---W---D---P---T--- 360  
Db 3480 KPTFRKGNLYEVIENDTITWDCGVTSRPLPSISWFRGDKPVVLYDRYSISPDGSHITI 3539  
QY 361 ---G---N---L---LA---S---C---S 368  
Db 3540 NKAKLSGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIVLECPIS 3599  
QY 369 ---D---M---D---D--- 371  
Db 3600 GIPOPDVITWKNMGDINWTSRVILAQNNETFGIENVQVTDQGRYCTATNRGKASHDF 3659  
QY 372 ---TL---K---I---W---S---M--- 378  
Db 3660 SLDVLSPPFDIHGTQPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGLDLD 3719  
QY 379 ---K---Q---D---D--- 381  
Db 3720 NVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPVVIDISRNDVQPV 3779  
QY 382 ---N---C---V---H---D---L---Q---Q---H--- 390  
Db 3780 AVNQPTIMRCVATGHPFPIKWLKNGKEVTDENIRIVEGQVQLIILRTSDSHAGKWSVCV 3839  
QY 391 ---N---KB---I---Y---TI---KW--- 399  
Db 3840 AENDAGVKELEWLDVFTPPVSVKSDNPIKALGETITLFCNAGSNPYPOLKWAQKGLI 3899  
QY 400 ---SP---T---G---P---GT---N--- 407  
Db 3900 FDSPDGARIKLGARLDIPLHKKTDVGYTCQALNAAGTSEASVSDVLVPPPEINRDGID 3959  
QY 408 ---N---P---P---P--- 409  
Db 3960 MSPRLPAQOSITLQCLAQKPVQMRWTLNGALTHTPGITVASDSTFIQINNVSLSDK 4019  
QY 410 ---N---A---NLN---LA---S--- 417  
Db 4020 GVYTCYAENVAGSDNLMYVDVQAPVISNGTKQVIEGELAVIECLVEGYPAPQVSWLR 4079  
QY 418 ---A---SPDS---:---:--- 422  
Db 4080 NGRVETGVQGVRYVTDGRMLTIEARSLDSGIYLCASNEAGSAQOAYTLEVLVSPKII 4139  
QY 423 ---T---VR---L---W---DV---:--- 429  
Db 4140 TSTPGVLTPSSGKFSPLPCAVRGVDPDIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4199  
QY 430 ---D---D---R---G---I---C---I--- 435  
Db 4200 RHLIYECTAKNADAGADTLEPPVQTVIAPKISTSGNRYINGSEGTETVIKCEISESEFS 4259  
QY 436 ---H---T---L---:---:---K 440  
Db 4260 WSKNGVPLPSNNLIFSEDYKLIKILSTRLSDOGEYSCTAANKAGNATQKTNLVGVAPK 4319  
QY 441 ---H---Q---E---P---V---Y---SVAF--- 450  
Db 4320 IMERPRTQVVHKGDQVTLWCEASGVQPAITWYKDNELLTNGVDETATTKKSVIFSGI 4379  
QY 451 SP---D---D---G---G--- 454  
Db 4380 SPSQAGVYTCKAENWVASTEDIDLIVMPPEVPPVPMNVSTNPTVFLSCNATGIPEP 4439  
QY 455 ---R---Y---LA---SGSF---D---K 464  
Db 4440 VISWMRDSNIAIQNNEKYQILGTLTALRNVLPPDDG-FYHCIAKSDAGQKIATRKILVVK 4498

QY 465 ---C---VHI---W---N--- 470  
Db 4499 PSDRPAPIWVECDKPKKTEYMDIRGDTDDNPQLPKWVEDSSLSNGSIAYRCMPGP 4558  
QY 471 ---T---O---T---GALV--- 477  
Db 4559 RSRTVLLHAPOPIVKKPKNTAAIGAIVELRCSAAGPPHPTITWAKDGKLIEDSKFEIA 4618  
QY 478 ---H---S--- 479  
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAVITCYE 4678  
QY 480 ---Y---RG---T---G---GI--- 486  
Db 4679 RNOAYSRGLTWEYNGVMPKNLAGIHFMMNGSLVILDTSSLKEGDLLEYTCVKVNRNRHS 4738  
QY 487 ---FE---V---C---W--- 491  
Db 4739 IPHLTSAFEGVPEVKTIKVEVNGSDSVLDCVTSDDLTHVVTWKNDQKMLDDDAIYV 4798  
QY 492 ---N---A---A---GD---K---V--- 498  
Db 4799 LPNNSLVLLNVEKYDEGVYKCVASNSIGKAFDDTQLNVYEGDFPLTGFEGSGINIDSS 4858  
QY 499 ---G---AS---:---:---A--- 502  
Db 4859 NAGGSRREAYKENEDASTTTITTSPTTTTETPLTTIIPALITLPAKQVPTDDYHE 4918  
QY 503 ---S---DG---SV---CV--- 509  
Db 4919 GSANDDGFPTQDLSLFEENPLHPISVNVNTDCAGTINENGDCVDKDGKTHNLKLTGE 4978  
QY 510 ---LD---:---:---L--- 512  
Db 4979 NHCPEGFAMNPRICEDLDCAFYQPCDFECINYDGGFCQCPLGYELAEBCRDVNEC 5038  
QY 513 ---R---K 514  
Db 5039 ESVRCEDGK 5047

RESULT 15  
US-10-369-493-6858  
; Sequence 6858, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6858  
; LENGTH: 5198  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-6858

Query Match 70.9%; Score 2595.3; DB 14; Length 5198;  
Best Local Similarity 9.0%; Pred. No. 2.3e-29;  
Matches 453; Conservative 50; Mismatches 9; Indels 4537; Gaps 389;

QY 1 M-----SI-----SS-----D-----E----- 7  
Db 1 MGRSPSWLYGVLLGILLLLATTCSSVNDKNDPTGKSSLAFFVDTGSMFDDLVQVREGAAK 60

QY 8 V-----N-----F--L--VY-----13  
Db 61 IFKTVMAQREKLIYNYIMVPHDPYLGEIINTTSTVFMRLSKVYVHGSGDCPEKTLTG 120  
QY 14 -----R--Y-----L--QE--SG--F-----SHSAP-T 27  
Db 121 ILKALQISLPSFIYFTDARSXDYHLEDEVLNTIOEKSSVVFVMTGCGNRTHPGFR 180  
QY 28 FG-I---KS-----H--ISO-----S-NI-----39  
Db 181 YEKIAAASFGQVPHLEKSDVSTVLEYVRHAKQKVHLMYEARERGTVSRNIPVDKHL 240  
QY 40 -----N-----GA-----LV--P-P-----46  
Db 241 ELTISLGDKDDNLDIVLRDPGRVTDKRLYSKEGGTIDLKNVKLIRLKDPSPGVWTV 300  
QY 47 -----A-----A-----LI--S-I-----52  
Db 301 NTSRLKHTIRVFGHGAUDFKYGFASRPLDRIELARPRVLNQDTYLLINMTGLIPPGTV 360  
QY 53 I-----Q-----KGL--Q-Y--VE--AE-63  
Db 361 GEIDLVDYHLSYKAVASPHRTNPNMYFAGPVPKGLFFVRVQGVDEDNIEFMRIAPT 420  
QY 64 V-----S-I-----N-----ED--GTLF-----73  
Db 421 AIGSVIVGGPRAPMSPHQEFVGRDLNLSCTVESASAYTIYWKGTGEDIIIGPLFYHNTD 480  
QY 74 -----D-----G-----R--P-I-----E-----79  
Db 481 TSVMTIPELSLKDAGEYECRVISSNNGNYSVKTRETRESPEIFGVRNVSVPLGEAAFLH 540  
QY 80 S-----L-----SL-I-----DA-----86  
Db 541 CSTRSAGEVEIRWTRGATVNGPNTERNPTNGTLKIHVTRADAGVYECMARNAGCMST 600  
QY 87 VM-----P-DV-----V-----92  
Db 601 RKMRLDIMEPPSVKVTQDVYFNMREGVNLSCAMGDPKPEVHWYFKGRHLLNDYKYVG 660  
QY 93 Q-----T--R--Q--QAYRD--KLA-----Q-----105  
Db 661 QDSKFLYIRDATHHDEGTVECRAMSQAQA-RDITDMLATPPKVEIIONKMMVGRDRV 719  
QY 106 -----Q-----Q-----A--A--A-----A-111  
Db 720 SPECKTIRGPKPKIRWFKNGKDLIKPDDYIKINEQLHMGAKEDAGAYSCVGNMAG 779  
QY 112 A-----A-----A-----A-----114  
Db 780 KDQVANLSVGRVPTIESPHTVRVNIERQVTIQLAVGIPPEIEWQKGNVLLATLNP 839  
QY 115 A-----A-----A-----ASQ-----120  
Db 840 RYTQADGNLLITDAIEDQOFTCIARNTYGOOSQSTILMTGLVSPVLGHVPPBEQLI 899  
QY 121 Q-----Q-----GS-----123  
Db 900 EGQDLTSCVVVLGTPKPSIVWIKDKPVEEGPTIKIEGGSLRLRGNPNKDEGKYTCI 959  
QY 124 A-----A-----K-----N-----GE--128  
Db 960 AVSPAGNSTLHINVQLIKPEFVYKPEGGIVFKPTISGMDKHAHVAVNSTHVDLGEFGA 1019  
QY 129 -----NT-----A-N-G-133  
Db 1020 IPCVSGTPTPIITWYLDGRPTFNSRDFVTADNTLIVRKADKSYGVVTCOATNSAGD 1079  
QY 134 E-----E-----N-136  
Db 1080 NEQKTTIRMTMISPGOSSFMNVDDLFTIPCDVYGPKPVTIWLDDKPTGEGVNE 1139  
QY 137 G-----AH-----TI-A-N-----N-----144

Db 1140 DGSLLTIPNVNEAHRGFTTCHQNAAGNDTRTTLTVHTTPTINAENQEKIALQNDIVLE 1199  
QY 145 -----HT--D-----MME-----V-----D-152  
Db 1200 CPAKALPPPPVRLWYEGEKIDSQLIPHTITREDGALVQLNQKLENTGVFVQVSNLAGEDS 1259  
QY 153 -----G-DV-----EIP-----P-----NK-AV-----163  
Db 1260 LSYTLTVHEKPKIISVPGVWDVVKGTIIEIPCRATGVPVIRTNKNGIDLKMDEKFFS 1319  
QY 164 V-----LR-----G-----H--E-----S-----170  
Db 1320 VDNLGLTRIYEADKNDIGNVNCVVTNEAGTSQMTTHVDVQEPPIILPSTQNTNNTAVVGDR 1379  
QY 171 -----E-----V--F-----I--C-----175  
Db 1380 VELKCYEASPPASVTWFRRGIAIGTDTKGYVYVESDGLTIVOSASVEDATITCKASNPA 1439  
QY 176 -----A-----176  
Db 1440 GKAEANLQVTVIASPDIKDPDVVTQESIKESHPPSLYCPVFSNPLPQISWYLNKPLIDD 1499  
QY 177 W-----178  
Db 1500 KTSWKTSDDKRLHVFKAKITDSGVYKCVARNAAGEGSKSFQVEVIVPLNLDSEYKVKV 1559  
QY 179 -----PVS-----181  
Db 1560 FAKGSEVTLGCPVSGFPVQINWVVDGTVVBPFGKYKGATLSNDGLTLHFDSSVKQEG 1619  
QY 182 -----D--L--LA-----S-----186  
Db 1620 NYHCVAQSKENILDDIVLSVLAVPIVGEDDNLEVLGKDISLSCOLQTESDDKTTFVNS 1679  
QY 187 --GS-----G-----DS-----T-----A-R-----194  
Db 1680 INGSERSDPNVQIIPSDGHRLYITDAKPENNGKMYCRVTNSAGKABRTLTLDVLEPPFV 1739  
QY 195 -----N-----N-----LS-----199  
Db 1740 EPVFEANQKLGNNPIILOQCVTGNPKPTVIWKIDGNDVDKSWLFDSELSLRIEKLTKG 1799  
QY 200 -----EN-----S-----S--T-----S-GS--T-207  
Db 1800 SAQISCTAENKAGTASRDPFIQNIAPTFKNEGDQETIFRESEITILDCPVSIGDQITW 1859  
QY 208 --Q-----L-VL--R-H-----C-----IR 216  
Db 1860 MKQGLPLTENDAIFTLDNTRLTILNARDHEDIYTCVANNTAGOVSKDFDVVVQVLPKIK 1919  
QY 217 -----EG-----G-----Q-DVP-----223  
Db 1920 NAVVTLEINEGEIILITCDAENPTPTAKWDFNQGLDLPKEAVFVNNHTVNNVNTKYHT 1979  
QY 224 -----SNK-----DV-----T-S-L--D-----232  
Db 1980 GUYKCVATNKVQAVKTIHVHTKPRFPESGLTESELVNLTRSTILECDVDDAIGVGIS 2039  
QY 233 W-N-----SE-G--TL--L-----A-T--GSY-----245  
Db 2040 WTVNGKFLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVTNEAGVATKTNLFVQVPP 2099  
QY 246 -----DG-----F-----248  
Db 2100 TIVNEGEYTVIENNSLPLCEVTGKPNPVVTTWKDGPVGLSKSVQLSEGGQFKIHA 2159  
QY 249 -----A-----R-----250  
Db 2160 EIAHKSICYMAKNDVGTAEISFDVDDIITRPMIQIKINIVTAIKGALPFPKPIDDDKN 2219  
QY 251 --IW-----T-----K-D-G-N-----257

Db 2220 FKGOIILWLNQYQPIDLEADARITRLSNDRLTILNVTEDEGOYSCRVKNKDAGENSDFD 2279  
QY 258 -----L-----A--S--TL-----QO--H----- 265  
Db 2280 KATVLVPPTIIMLDKDKNTAVEHSTVTLSCPATGKPEPDITWFKDGEAIHENTADIIP 2339  
QY 266 -----G-----HT-----N--E-----VN-----A--I-- 355  
Db 3420 PTVTWKRGDDAVPNDSSRTIVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVI 3479  
QY 356 -----K-----N-----L-----LA-----S--C--S 368  
Db 3480 KPTFVRKGNLYEVIENDITMDCGVTSRPLPSISWFRGDKPVYLDVRSISPDSGSHITI 3539  
QY 361 -----G-----N-----L-----LA-----S--C--S 368  
Db 3540 NKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPIS 3599  
QY 369 -----D--M-----D-----LA-----S--C--S 371  
Db 3600 GIPQPDVITWKGMDINMTDSRVILAONNETFGIENVQVTDQGRYTCTATNRGGKASHDF 3659  
QY 372 -----TL-----K-----I-----W-----S--M-- 378  
Db 3660 SLDLVSPBFDIHGTQPTIKREGDTITLCPKLAEDIAQVMDVSWTKDSRALDGLD 3719  
QY 379 -----K--Q-----D-----LA-----S--C--S 381  
Db 3720 NVDISDDGRKLITISQASLENAGLYTCIALNRAGEASLEFKVEILSPPIDISRNDVQPV 3779  
QY 382 --N-----C-V--H-----D--L--O--O--O-----H-- 390  
Db 3780 AVNQPTIMECATGHPFPIKMLKNGKEVTDENIRIVEQGVQLLRTDSDHAGKWSV 3839  
QY 391 --N-----KE-----I--Y--TI-----KW----- 399  
Db 3840 AENDAGKELEMLVDVFTPPVSVKSDNPIKALGETITLFCNAGSNPYQLKWKAGGSLI 3899  
QY 400 --SP-----T--G--P-----GT-----N----- 407  
Db 3900 FDSPDGARISLKGARLDIPHLKKTVDGYTCOALNAAGTSEASVSDVILVPEINRDGID 3959  
QY 408 -----N-----P----- 409  
Db 3960 MSPRLPAQSLTLQCLAQKPVQMRWLTNGTALHTSTFGITVASDSTFIQINNVSLSK 4019  
QY 410 -----N-A--NLM-----LA-----S-- 417  
Db 4020 GYVTCYAENVAGSNDLMTNVVDVVOAPVISNGGKQVIEGELAVIECLVEGYPAPQVSM 4079  
QY 418 -----A-SFDS----- 422  
Db 4080 NGRVETGVQVRYVTDGRMLTIIERSLDSGIYLCSATNEAGSAQAYTLEVLSPKII 4139  
QY 423 -----T-----VR-L-----W-----DV----- 429  
Db 4140 TSTPGVLTSPSSGSKFSLPCAVRGYDPDIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4199  
QY 430 -----D-----R-----G--I--C--I----- 435  
Db 4200 RHLIYECTAKNDAGADTLEFPVQTVIAPKISTSGNRYINGSEGTETVIKCEIESESEFS 4259  
QY 436 -----H-----T--L-----K-- 440  
Db 4260 WSKNGVLLPNNLIFSEYDKLILSTLSDQGEYSCTAANKAGNATQKTNLVGVAPK 4319  
QY 441 -----H--Q-----E-----P-V--Y-----SVAF-- 450  
Db 4320 IMERPRTQVHKGDQVTLWCEASGVQPAITWYKDNELLTNTGVDETATTKKSVIFSSI 4379  
QY 451 SP-----D-----G----- 454  
Db 4380 SPSQAGVYTCKAENWASTEEDIDLIVMIPPEVPERMNVSTNPRQTVFLSCNATGIP 4439

Db 2220 FKGOIILWLNQYQPIDLEADARITRLSNDRLTILNVTEDEGOYSCRVKNKDAGENSDFD 2279  
QY 258 -----L-----A--S--TL-----QO--H----- 265  
Db 2280 KATVLVPPTIIMLDKDKNTAVEHSTVTLSCPATGKPEPDITWFKDGEAIHENTADIIP 2339  
QY 266 -----G-----HT-----N--E-----VN-----A--I-- 355  
Db 3420 PTVTWKRGDDAVPNDSSRTIVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVI 3479  
QY 356 -----K-----N-----L-----LA-----S--C--S 368  
Db 3480 KPTFVRKGNLYEVIENDITMDCGVTSRPLPSISWFRGDKPVYLDVRSISPDSGSHITI 3539  
QY 361 -----G-----N-----L-----LA-----S--C--S 368  
Db 3540 NKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPIS 3599  
QY 369 -----D--M-----D-----LA-----S--C--S 371  
Db 3600 GIPQPDVITWKGMDINMTDSRVILAONNETFGIENVQVTDQGRYTCTATNRGGKASHDF 3659  
QY 372 -----TL-----K-----I-----W-----S--M-- 378  
Db 3660 SLDLVSPBFDIHGTQPTIKREGDTITLCPKLAEDIAQVMDVSWTKDSRALDGLD 3719  
QY 379 -----K--Q-----D-----LA-----S--C--S 381  
Db 3720 NVDISDDGRKLITISQASLENAGLYTCIALNRAGEASLEFKVEILSPPIDISRNDVQPV 3779  
QY 382 --N-----C-V--H-----D--L--O--O--O-----H-- 390  
Db 3780 AVNQPTIMECATGHPFPIKMLKNGKEVTDENIRIVEQGVQLLRTDSDHAGKWSV 3839  
QY 391 --N-----KE-----I--Y--TI-----KW----- 399  
Db 3840 AENDAGKELEMLVDVFTPPVSVKSDNPIKALGETITLFCNAGSNPYQLKWKAGGSLI 3899  
QY 400 --SP-----T--G--P-----GT-----N----- 407  
Db 3900 FDSPDGARISLKGARLDIPHLKKTVDGYTCOALNAAGTSEASVSDVILVPEINRDGID 3959  
QY 408 -----N-----P----- 409  
Db 3960 MSPRLPAQSLTLQCLAQKPVQMRWLTNGTALHTSTFGITVASDSTFIQINNVSLSK 4019  
QY 410 -----N-A--NLM-----LA-----S-- 417  
Db 4020 GYVTCYAENVAGSNDLMTNVVDVVOAPVISNGGKQVIEGELAVIECLVEGYPAPQVSM 4079  
QY 418 -----A-SFDS----- 422  
Db 4080 NGRVETGVQVRYVTDGRMLTIIERSLDSGIYLCSATNEAGSAQAYTLEVLSPKII 4139  
QY 423 -----T-----VR-L-----W-----DV----- 429  
Db 4140 TSTPGVLTSPSSGSKFSLPCAVRGYDPDIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4199  
QY 430 -----D-----R-----G--I--C--I----- 435  
Db 4200 RHLIYECTAKNDAGADTLEFPVQTVIAPKISTSGNRYINGSEGTETVIKCEIESESEFS 4259  
QY 436 -----H-----T--L-----K-- 440  
Db 4260 WSKNGVLLPNNLIFSEYDKLILSTLSDQGEYSCTAANKAGNATQKTNLVGVAPK 4319  
QY 441 -----H--Q-----E-----P-V--Y-----SVAF-- 450  
Db 4320 IMERPRTQVHKGDQVTLWCEASGVQPAITWYKDNELLTNTGVDETATTKKSVIFSSI 4379  
QY 451 SP-----D-----G----- 454  
Db 4380 SPSQAGVYTCKAENWASTEEDIDLIVMIPPEVPERMNVSTNPRQTVFLSCNATGIP 4439

QY 455 -----R-----Y-----LA-----SGSF-----D-----K 464  
Db 4440 VISWEDSNIAIONNEKYQILGTTLAIRNVLPPDDG-FYHCIAKSDAGQKIATRKLIYNK 4498  
QY 465 -----C-----VHI-----W-----N-----470  
Db 4499 PSDRPAPIWVECKEKGPKKTEYMDIRGTPDDNPOLLPWKOVESLNGSIAYRCMPGP 4558  
QY 471 -----T-----Q-----T-----CALV-----477  
Db 4559 RSRRTVLLHAAQPIVFKPKNTAAIGAIIVELRCSAAGPPHPHTITWAKDKLIEDSKFEIA 4618  
QY 478 -----H-----S-----479  
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSAPINVDNNILPTPKPSSNQKNVAVITCYE 4678  
QY 480 -----Y-RG-T-----G-----GI-----486  
Db 4679 RNQAYSRLTWEYNGVMPKPNLAGIHFMNNGSLVILDTSSLKEGDLLEYTCVKNRRRHS 4738  
QY 487 -----FE-----V-C-----W-----491  
Db 4739 IPHLSAFEGVEVKTIQKVEVNGDSVVLDCVTSOPLTHVWTKNDQKMLDDDAIYV 4798  
QY 492 -----N-----A-----A-----GD-----K-----V-----498  
Db 4799 LPNNSLVLLNVEKYDGVYKCVASNSIGKAFDDTQLNVYEGDFLPLTGFECSGINIDSS 4858  
QY 499 -----G-----AS-----S-----492  
Db 4859 NAGSSRRRAYKENEDASTTTTTSPTTTTTEPLTTTIIIPALITLPAKQYPTDDYHE 4918  
QY 503 -----S-----DG-----SV-----CV-----509  
Db 4919 GSANDGFGFTQDSLFEFNPPLHPEISVUNTDCAGTINENGDCVDKDKTNLKITLGE 4978  
QY 510 -----LD-----L-----512  
Db 4979 NHCPEGFAMNPHTRICEDLDECAFYQPCDFECINYDGFQCNCPGLGYELAEBCRDVNEC 5038  
QY 513 -----R-----K 514  
Db 5039 ESVRCEDGK 5047

Search completed: January 3, 2005, 15:47:16  
Job time : 136 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:09:41 ; Search time 21.6667 Seconds  
(without alignments)  
2282.558 Million cell updates/sec

Title: US-09-987-701-2

Perfect score: 3659

Sequence: 1 MSISSDEVNFLVRYLQESG.....GDKVGASGDSVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2596.6	71.0	5175	T20992	hypothetical prote
2	2595.3	70.9	5198	T43290	hemacentin precurs
3	2594.7	70.9	7962	I38346	elastic titin - hu
4	2585.4	70.7	26926	I38344	titin, cardiac mus
5	2583.7	70.6	10797	T30192	probable peptide s
6	2580.5	70.5	4544	S02392	alpha-2-macroglobu
7	2580.2	70.5	6658	T13331	projectin - fruit
8	2578.7	70.5	7829	T15789	hypothetical prote
9	2574.9	70.4	4660	T42737	gp330 protein prec
10	2572.2	70.3	6831	T48852	protein unc-22 [im
11	2572.2	70.3	6839	S57242	twitchin [similar
12	2572.2	70.3	7160	T27935	hypothetical prote
13	2571.1	70.3	5825	T12117	polyprotein - fava
14	2569.5	70.2	4545	T52511	alpha-2-macroglobu
15	2568.8	70.2	3461	S58870	reelin precursor -
16	2567.9	70.2	5232	T45086	HC-toxin synthetas
17	2561.3	70.0	4543	A53102	alpha-2-macroglobu
18	2560.7	70.0	5376	T42215	zonadhesin - mouse
19	2559.8	70.0	6805	T20901	titin - rabbit (fr
20	2556.6	69.9	8243	T31307	type I fatty acid
21	2554.4	69.8	4753	A47437	LDL-receptor-relat
22	2543.3	69.5	6669	S55024	nebulin, skeletal
23	2539.8	69.4	15281	S41309	cyclosporin synthe
24	2539	69.4	4767	T31345	hypothetical prote
25	2537.7	69.4	4836	T14346	herc2 protein - mo
26	2537.5	69.3	9376	T14593	syngomycin synth
27	2536.4	69.3	4302	A38971	polycystic kidney
28	2530.8	69.2	3623	T09456	intrinsic factor-B
29	2529	69.1	6642	T29757	protein UNC-89 - C

30	2526.3	69.0	4861	2	S71752	giant protein p619
31	2526.1	69.0	3623	2	T08618	intrinsic factor-B
32	2522.8	68.9	5138	2	B96695	hypothetical prote
33	2520.1	68.9	4930	2	E69679	polyketide synthet
34	2519.1	68.8	4464	2	D87755	protein T21E12.4 [
35	2519	68.8	4639	1	A54794	dynemin heavy chain
36	2514.7	68.7	4644	1	A38905	dynemin heavy chain
37	2512.8	68.7	4568	2	T08030	dynemin beta heavy
38	2512	68.7	4344	1	A53489	hypothetical 527K
39	2509.4	68.6	4572	2	S57908	hypothetical synth
40	2508.6	68.6	5369	2	T44807	mycosubtilin synth
41	2504.6	68.5	4447	2	A69679	polyketide contain
42	2504.2	68.4	4845	2	T31067	Bir repeat contain
43	2502.3	68.4	3871	2	T22812	hypothetical prote
44	2502.1	68.4	3856	2	T51174	ataxia-telangiecta
45	2497.7	68.3	4391	2	A38096	perlecan precursor

ALIGNMENTS

RESULT 1

T20992  
hypothetical protein F15G9.4a - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T20992; T24733  
R:Sulston, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19355  
A:Accession: T20992  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-5175 <W1L>  
A:Cross-references: UNIPROT:Q810L3; EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a  
R:Kershaw, J.  
submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19929  
A:Accession: T24733  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-5175 <W12>  
A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a  
A:Experimental source: clone T09B9  
A:Gene: CESP:F15G9.4a  
A:Map position: X  
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3; 1251/2; 2593/3; 2699/3; 2759/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5077/1

Query Match 71.0%; Score 2596.6; DB 2; Length 5175;  
Best Local Similarity 9.0%; Pred. No. 6.6e-53;  
Matches 454; Conservative 48; Mismatches 10; Indels 4514; Gaps 390;

QY	1	M-----SI-----SS-----D-----E-----	7
Db	1	MGRSPSWLYGVLLGLLLATTCSSVNDKNDPTGKSLAFVFDITGSMFDDLVVVRGA	60
QY	8	-----V-----N-----F-----L-----V-----	13
Db	61	IFKTVMAQREKLYINWVPHDPYLGEIINTDSTYFMRQLSKVYVHGGDCPEKLT	120
QY	14	-----R-----Y-----L-----QE-----SG-----F-----	SHAF-T 27
Db	121	ILKALQISLPSSFIYVFTDARSKDVLHEDVLTIOEQSSVVFVMTGCGNRTHPGFRT	180
QY	28	-----FG-I-----KS-----H-ISO-----S-NI-----	39
Db	181	YEKIAAASFGQVFHLEKSDVSTVLEYRHAHVQKKVHLMYEAEREGTVSRNIPVDK	240
QY	40	-----N-----GA-----LV-----P-P-----	46

Db	241	ELTISLGDKDDSDNLDIVLRDPEGRVTDKRLYSKEGGTIDLDKNVXKILRLKDPSPGVWTV	300
Qy	47	-----A-----A-----A-----LI--S--I-----	52
Db	301	NTNSRLKHTIRVFHGAVDKFKYGFASRPLDRIBELARPRPVLNQDTYLLINMTGLIPPGTV	360
Qy	53	-----Q-----KGL-----O--Y-----VB-----AE--63	
Db	361	GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPPVPPKGLFFVRVQGVDEDNIEFMRIAPT	420
Qy	64	-----V-----S--I-----N-----ED-----GTLF-----	73
Db	421	AIGSVIUGGPRAFMSPHQEFVGRDLNLSCTVESASAYTIYWKYTGEDIIGGLFLFYHNTD	480
Qy	74	-----D-----G-----R--P--I-----E-----	79
Db	481	TSVMTIPELSKDAGEYECRVISNNGNYSVKTRVETRESPEIFGVNRVNSVPLGEAAFLH	540
Qy	80	-----S-----L--I-----DA-----	86
Db	541	CSTRSAGEVEIRWTRGATVNGPNTERNPTNGLKIHVTRADAGVYECMARNAGMST	600
Qy	87	-----VM-----P--DV-----A-----V--92	
Db	601	RKRLDIMEPPSVKVTPODYVFNMRBGVNLSCAMGDKPKPEVHWYFKGRHLLNDYKYQVG	660
Qy	93	-----T--R--Q--QAYRD--KLA-----Q-----	105
Db	661	QDSKFLYIRDATHDEGTVECRAMSQAQA-RDITDMLATPPKVBIIQNMWVGRDRV	719
Qy	106	-----Q-----Q-----A-----A-----A--111	
Db	720	SPECKTIRGPKPKIRWFKNGKDLIKPDDYIKINEGOLHMGAKDEBDAGAYSCVENMAG	779
Qy	112	-----A-----A-----A-----A-----A--114	
Db	780	KDQVANLSVGRVPTIIESPHTVRVNIQROVTQLCLAVGIPPEIEWQKGNVLLATLNP	839
Qy	115	-----A-----A-----A-----ASQ-----	120
Db	840	RYQLADGNLLITDAIEDQOFTCIARNTYGOQSOSTLLMTGLVSPVLGHVPPBEQLI	899
Qy	121	-----Q-----GS-----	123
Db	900	EGQDLTSCVVLGTPKPSIWIKDDKPVEBEGTIKIEGGSLRLRGNPKDEGKYTCI	959
Qy	124	-----A-----K-----N-----GE--128	
Db	960	AVSPAGNSTLHINVQLIKKPEFYVKEGGIVFKPTISGMDKXHVAVVNSTHVDLDGEFA	1019
Qy	129	-----NT-----A--N--G--133	
Db	1020	IPCWSGTPPIITWYLDGRIPTPNSRDFVTADNTLIVRKADKSYGVVTCQATNSAGD	1079
Qy	134	-----E-----E--N--136	
Db	1080	NEQKTIIRIMNTMISPGQSSFNWVDDLETPICDVYGDPKPVITWLLDDKPTGVVNE	1139
Qy	137	-----AH-----TI--A--N-----N-----144	
Db	1140	DGSLTIPNVNEAHRGFTCHAQNAAGNDRTVTLTHTVHTPTINAEQEKIALQNDIVLE	1199
Qy	145	-----HT--D-----MMB-----V-----D--152	
Db	1200	CPAKLPPVRLWYBEGEKIDSQI1PHTIREDGALVQNVKLENTGVFVQVSNLAGEDS	1259
Qy	153	-----G--DV-----EIP-----P-----NK--AV-----163	
Db	1260	LSYLTVHEKPIISEVGVVDVVVKFTIEIPCRATGVPEVIRTNWNGIDLKMDKFKFS	1319
Qy	164	V-----LR-----G-----H-----E-----S-----170	
Db	1320	VONLGLRIYEADKNDIGNYNCVVTNEAGTSQMTTHVDVQEPFIILPSTQNTTAVVGDR	1379
Qy	171	-----E-----V--F-----I--C-----175	
Db	1380	VELKCYVEASPPASVTFRRGIAIGTDTKGYVVESDGLTVIOSASVEDATITYTKASNPA	1439
Qy	176	-----A-----	176
Db	1440	GKAEANLQVTVIASPDIKDPDVVTQESIKESHPPFSLYCPVFSNPLQISWYLNKDLIDD	1499
Qy	177	-----W-----N-----178	
Db	1500	KTSWKTSDDKRXLHVFKAKITDSGVYKCVARNAAGBSKSFQVEVIVPLNLDSEYKXKV	1559
Qy	179	-----PVS-----181	
Db	1560	FAKEGEVTLGCPVSGFPVQINWVVDGTVVBPFGKYKGATILNDGLTLHFDSSVSKQEG	1619
Qy	182	-----D--L--LA-----S--186	
Db	1620	NTHCVAQSKGNILDDIVELSVLAIVPIVGEDDNLLEVLGKDISLSCDLQTESDDKTTFVMS	1679
Qy	187	-----GS-----G-----DS-----T-----A--R-----194	
Db	1680	INGSSEDRPDNVQIPSDGHRLYITDAKPENNNGKYMCRVTNSAGKABRTILTLVLEPPFV	1739
Qy	195	-----IW-----N-----LS-----199	
Db	1740	EPVFEANQKIGNNPIILOQCVTGNPKPTVIWKIDGNDVDKMWLFDESLSLRLEKLTGK	1799
Qy	200	-----EN-----S--T-----S--GS--T--207	
Db	1800	SAQISCTAENKAGTASRDFIQAIAPTFKNEGDQETIPRESEITILDCPVSLGDFQITW	1859
Qy	208	-----Q-----L--VL--R--H-----C-----IR--216	
Db	1860	MKQGLPLTENDAIFTLDNTRLTILNARDHEDITYTCVANNTAGQVSKDFVVVQVLPKIK	1919
Qy	217	-----EG-----G-----O--DVP-----223	
Db	1920	NAVVTLEINEGEBEILTCDAEGNPPTAKWDFNQGLPKAEAVFVNNHTVVNNVTKYHT	1979
Qy	224	-----SNK-----DV-----T--S--L-----D-----232	
Db	1980	GVYKCATNKVQAVKTIINVHVTKRPFESGLTESELTVNLTFESITLECDVDDAIGVGIS	2039
Qy	233	W--N-----SE--G--TL-----L-----A--T-----GSY-----245	
Db	2040	WTWNGKFLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVTNEAGVATKTNFLVQVPP	2099
Qy	246	-----DG-----F-----248	
Db	2100	TIVNEGEGVTVIENNSLVLPCVTKPNPVVTWTKDGRPVGDLKSVQVLSGGOQPKIVHA	2159
Qy	249	-----A-----R-----250	
Db	2160	ETAHKGSYICMAKNDVGTAEISFDVDIITRPMIQKIKNIVTAIKGALPFKPCPIDDDKN	2219
Qy	251	-----IW-----T-----K--D--G--N-----257	
Db	2220	FKQIILWLNQYPIIDLEAEDARITRLSNDRLTLILNVTEDEBQYSCRVKNDAGENSFDF	2279
Qy	258	-----L-----A--S--TL-----GO--H-----265	
Db	2280	KATVLPPTIIMLDKDKNKTAVEHSTVTLSCPATGKPEPDITWFKDGEAIHINIADIIP	2339
Qy	266	-----K--G--P-----I--269	
Db	2340	NGELNQNQLKTIKEGDAGKYTCADNSAGSVEQDVNVNVTIPKIEKDIPSDYESQQ	2399
Qy	266	-----K--G--P-----I--269	
Db	2400	NERVVISCPVYARPPAKITWLKAGKPLQSKDFVKTSANGQKLYLFKLRETDSSKYTCIAT	2459

QY 270 -----P-----A-----L-----K-----W-----N----- 275  
Db 2460 NEAGTDKDFKVMVAPSFDEPNVIRITVNSGNPSTLHCPAKGSPSPITITWLKGNAI 2519  
QY 276 -----K----- 276  
Db 2520 EPNDRYVFFDAGRQLQTSKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPPVIDGERREA 2579  
QY 277 -----K-G-N-----FI-----LS----- 283  
Db 2580 VAVIEGSELSFCDNSSTGVVDVWQDGLTINQDTLRGDSFIQIPSSGKKMSFLSARKSD 2639  
QY 284 -----AG-----V-----DK-T-----TI-----I----- 292  
Db 2640 SGRYTCIVRNPAAGEARKLPDFAVNDPPSISDELSSANIQTIVPYVPVEINCVVSGSPHPK 2699  
QY 293 --W-----D--A-----H-----T-----G----- 298  
Db 2700 VYMLFDDKPLEPDSAAVELTNNGETLKIVRSQVEHAGTYTCEAQNNGVARKDFLVRVTA 2759  
QY 299 -----E-----AK-----Q----- 302  
Db 2760 PPHFEKERBEVWARVGDMLLTACNASSVFLSSVYMAHDESVQNGVITSKYAANEKTLN 2819  
QY 303 -----Q-----F----- 304  
Db 2820 VTNIQLDDEGYFYCTAVNEAGITKPKFLIVIEFPYFLDOOKLYPIILGKRLTDCSATG 2879  
QY 305 --P-----F----- 306  
Db 2880 TPEPTILFMKDGRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGRSKDMWVEV 2939  
QY 307 -----H-----S----- 308  
Db 2940 LLPPKLSKEWINVEVOAGDPLTLECPIDETSGVHITWSRQFGKQDGMRAQSSDKSL 2999  
QY 309 -----A-P--A--L-----D-----V-D----- 315  
Db 3000 YIMQATPEDADSCTIANDAGAEAVQVVTNTPPKIFGDSPTTEIVADTLEIPCRT 3059  
QY 316 -----W-----Q----- 317  
Db 3060 EGIPPEISWFLDGKPILEMPGVYTKQDLSLRIDNIKPNQEGRYTCVAENKAGRAEQDT 3119  
QY 318 -----S-----N-----T-----F----- 322  
Db 3120 YVEISEPPRVVMASEVMRVVEGRQTTIRCEVFGNPEPVNWLKDGEPYTSLLQFSTKLS 3179  
QY 323 -----A-----SC-----S-----TD-----M--C-- 330  
Db 3180 YLHLRETTLDGGTYTCIATNKAGESQTTIDVEVLVPPRIEDEERVLQKEGNTYVWHQ 3239  
QY 331 -----I-----HV-----C-----K 335  
Db 3240 VTGRPVYVTKRNGKEIEQFNPLHNRATRADEKGYSCIASNEAGTAVADFLIDVFTK 3299  
QY 336 -----L--GQ-----D-----RP-----I-K- 343  
Db 3300 PTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGRPPFNMNDIILSPRGDTLMILKA 3359  
QY 344 -----T-----P-----Q----- 346  
Db 3360 QRPDGLGYTCVATNSYGDSEQDFKNVYTKPYIDETIDQTPKAVAGSILKCPVLGNFT 3419  
QY 347 -----G-----HT-----N-----E-----VN-----A--I-- 355  
Db 3420 PTVTKRGDDAVPNDSSRTITVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVIG 3479  
QY 356 -----K-----W-----D-----P-----T----- 360  
Db 3480 KPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVYLYDRYSISPDGSHITI 3539

QY 361 -----G-----N-----L-----LA-----S-----C--S 368  
Db 3540 NKAKLSDGKGYICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIYLECPIS 3599  
QY 369 -----D-----D-M----- 371  
Db 3600 GIPQPDVITWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGGKASHDF 3659  
QY 372 -----TL-----K-----I-----W-----S-----M-- 378  
Db 3660 SLDLVSPPFEDIHGTOPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGLDLD 3719  
QY 379 -----K-----Q-----D----- 381  
Db 3720 NVDISDDGRKLITISQASLENAGLYTCIALNRAGEASLEFKVEILSPVIDISRNDVQPV 3779  
QY 382 --N-----C-V--H-----D-----L-----Q-----Q-----H----- 390  
Db 3780 AVNQPTIMRCAVTGHPFSPSIKMLKNGKEVTDDEINIRIVEQGQVLQILRTDSDHAGKMSCV 3839  
QY 391 --N-----KE-----I--Y--TI-----KW----- 399  
Db 3840 AENDAGVKEMLVDVFTPPVSVKSDNFIKALGETITLFCNASGNPYPOLKWAKGSLI 3899  
QY 400 --SP-----T--G-----P-----GT-----N----- 407  
Db 3900 FDSPDGARISLKGARLDIPLHKKTDVGDYTCQALNAAAGTSEASVSDVLVPPPEINRDGID 3959  
QY 408 -----N-----P----- 409  
Db 3960 MSPRLPAQOSLTLOCLAQKVPQMRWTLNGTALTHTPGITVASDSTFIQINNVSLSDK 4019  
QY 410 -----N-A--NLM-----LA-----S----- 417  
Db 4020 GVYTCVAENVAGSDNLMVNVVQAPVINSNGTKQVIEGELAVIECLVEGYPAPQVSWLR 4079  
QY 418 -----A-SFDS----- 422  
Db 4080 NGRVETGVQGVYVTDGRMLTIEARSLSDSGIYLCSATNEAGSAQAQAYTLEVLVSPKII 4139  
QY 423 -----T-----VR-L-----W-----DV----- 429  
Db 4140 TSTPGVLTPSSGSKFSLPCAVRGYPDPPIISWTLNGNDIKDGENGTIGADGTLHBAEE 4199  
QY 430 -----D-----R-----G-----I-C-I----- 435  
Db 4200 RHLIVECTAKNDAGADTLEFPVQTVIAPKISTSGNRYINGSEGTETVIKEIESESSEFS 4259  
QY 436 -----H-----T--L-----K 440  
Db 4260 WSKNGVPLLPNNLIFSEDYKLIKILSTRLSDOGEYSCTAANKAGNATQKTNLNVGVAPK 4319  
QY 441 -----H--Q-----E-----P-V--Y-----SVAF----- 450  
Db 4320 IMERPRTQVHVKGQDQVTLWCEASGVQPAPITWYKONELLTNTGVDETATTTKKKSVIFSSI 4379  
QY 451 SP-----D-----G----- 454  
Db 4380 SPSQAGVYTCBAENVVASTEEDIDLIVMTPPEVPMNVSTNPROTVFLSCNATGIPEP 4439  
QY 455 -----R-----Y-----LA-----SGSF-----D-----K 464  
Db 4440 VISWNRDSNIAIQNNEKYQILGTLTILAIRNVLPDDG--FYHCIAKSAGQKIATKILVNK 4498  
QY 465 -----C-----VHI-----W-----N----- 470  
Db 4499 PSDRPAPIWVECEKPKPKTEYMDRGDTPODNPOLLPKDVEDSSLSNGSIAYRCMPGP 4558  
QY 471 -----T-----Q-----T--CA-----LV----- 477  
Db 4559 RSRVTLLHAAPQFIVKPKNTAAIGAIVELRCSAAGPPHPTITWAKOKLIEDSKFEIA 4618  
QY 478 --H-----S----- 479

Db 4619 YSHLKVTILNSTSDSEYTCMAQNSVGSSTVSAPFINVDNNILPTPKPSNQKNVAVITCYE 4678  
Qy 480 ---Y-RG-T---G---GI--- 486  
Db 4679 RNQAYSRGLTWEYNGVMPKPLAGIHFWNGLSVILDTSSLKEGDLLEYLTKVNRNRHS 4738  
Qy 487 ---FE---V-C---W-N---A--- 493  
Db 4739 IPHLTSAFEGVEPKTKVEVNGDSVLVDCEVTSPLTHVVTNKDKMLDDDAIYV 4798  
Qy 494 ---A---G---D---KV---G---AS--- 501  
Db 4799 LPNNSVLILNVEKYDEGVKCVASNSIGKAFDDQLNLYGSSRRREAYKKNEDASTTTI 4858  
Qy 502 ---A---S---DG--- 505  
Db 4859 TTTSPTTTTTETPLTTIIPALITLPAKQYPTDDYHEGSANDDGFPTTQDSLFEFNPL 4918  
Qy 506 ---SV---CV---L---R---K 514  
Db 4919 HPEISVNTDCAGTINENGDCVDKGTNHLKILTGENHCPEGFPMNPHTRICEDLDECA 4978  
Qy 512 ---L---R---K 514  
Db 4979 FYQPCDFECINYDGGFCNCPLGYELAEAGCRDWNCECVRCEDGK 5024

RESULT 2  
T43290  
hemictentin precursor - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 11-Jan-2000 #sequence revision 11-Jan-2000 #text\_change 09-Jul-2004  
C:Accession: T43290; T20993; T24734  
R:Voegel, B.E.; Hedgecock, E.M.  
A:Submitted to the EMBL Data Library, June 1998  
A:Description: Hemictentin is required for hemidesmosome mediated cell adhesion and germ-  
A:Reference number: 222396  
A:Accession: T43290  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-5198 <WOG>  
A:Cross-references: UNIPROT:O76518; EMBL:AF074901; PIDN:AAC36792.1  
R:Sulston, J.  
A:Submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19355  
A:Accession: T20993  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5198 <WIL>  
A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CBSP:F15G9.4b  
A:Experimental source: clone F15G9  
R:Kershaw, J.  
A:Submitted to the EMBL Data Library, December 1994  
A:Reference number: Z19929  
A:Accession: T24734  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-5198 <W12>  
A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CBSP:F15G9.4b  
A:Experimental source: clone T09B9  
C:Genetics:  
A:Gene: him-4; F15G9.4b  
A:Map position: X  
A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;  
A:Exons: 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;  
A:Indels: 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/1

Query Match 70.9%; Score 2595.3; DB 2; Length 5198;  
Best Local Similarity 9.0%; Pred. No. 7.2e-53;  
Matches 453; Conservative 50; Mismatches 9; Indels 4537; Gaps 389;

Qy 1 M-----SI-----SS-----D-----B----- 7

Db 1 MGRSPWLYGVGLGLLLIATTCSSVNDKNDPTGKSLAFVFDITGSMFDDLVQREGAAK 60  
Qy 8 ---V---N---F---L---VY--- 13  
Db 61 IFKTVMAQREKLYNYIMVFFHDPYLGEINTTSTDYFMRQLSKVYVHGSGDCPEKTLTG 120  
Qy 14 ---R---Y---L---QF---SG---F---SHSAF-T 27  
Db 121 ILKALQISLPSSFIYFTDARSXDYHLEDEVLTIOEKQSSVVVFMVTGCGNRTHFGFRT 180  
Qy 28 ---FG-I---KS---H-ISO---S-NI--- 39  
Db 181 YEKIAAASFGQFHLKESDVSTVLEVRHAKQKVKHLMYEAERGGTVSRNIPVDKHL 240  
Qy 40 ---N---GA---LV---P-P--- 46  
Db 241 ELTISLSDGKDDNDLDIVLRDPEGRVTKRLYSKEGGTIDLKNVKILRLKDPSPGVWTV 300  
Qy 47 ---A---LI---S-I--- 52  
Db 301 NNTSLKHTIRVFGHGAVDKYGFAFRPLDRIELARPRVPLNQDTYLLINMTGLIPGTV 360  
Qy 53 ---I---Q---KGL---Q-Y---VE---AE- 63  
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGFPVPPKGLFFVRVQGYDEDNYEFMIAPT 420  
Qy 64 ---V---S-I---N---ED---GTLF--- 73  
Db 421 AIGSVIVGGPRAFMSPIHQEFVGRDLNLSCTVESASAYTIYVVKTGEDIIGGLFLYHNTD 480  
Qy 74 ---D---G---R---P-I---E--- 79  
Db 481 TSVWTTPELSLKDAGEYECRVISNNGNYSVKTRVETRESPEIFGVNRVNSVPLGEAFLH 540  
Qy 80 ---S---L---SL-I---DA--- 86  
Db 541 CSTRSAGEVEIRWTRYGAIVFNGPNTERTNGTLKIHVTRADAGVYECMARNAGMST 600  
Qy 87 ---VM---P-DV---V- 92  
Db 601 RKMRLDIMEPPSVKVTQDVFYFMRBEGVNSCEAMGDPKPEVHWYFKGRHLLNDYKYQVG 660  
Qy 93 Q---T---R---Q---QAYRD---KLA---Q--- 105  
Db 661 QDSKELYIRDATHDEGTYECRAMSQAGQA-RDITDMLATPKVEIIONKMMVGRDRV 719  
Qy 106 ---Q---Q---A---A---A---A- 111  
Db 720 SPECKTIRGPHPKIRWPKNGKDLIKPDDYIKINEQLHMGAKDEADAGAYSCVGENMAG 779  
Qy 112 ---A---A---A---A---A---A- 114  
Db 780 KDQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEPEIEWQKGNVLLATNPN 839  
Qy 115 ---A---A---A---ASQ--- 120  
Db 840 RYQLADGNLLITDAQIEDQGQFTCIARTYQQSQSSTLMTGLVSPVLGHVPPEEQLI 899  
Qy 121 ---Q---GS--- 123  
Db 900 EGQDLTLCVVVLGTPKPSIVMIKDDKPVIEGPTTIKIEGGGSLRLRGGNPKDEGKYTCI 959  
Qy 124 ---A---K---N---GE--- 128  
Db 960 AVSPAGNSTLHINVLIIKKPEFVYKPEGGIVFKPTISGMDEKXHVAVNSTHVDLGEQFA 1019  
Qy 129 ---NT---A-N---G- 133  
Db 1020 IPCVSGTTPPIITWLDGRPTTPNSRDTVTADNTLIVRKADKSYGVYTCQAINSGD 1079  
Qy 134 -E-----N-----E----- 136

Db 1080 NEQKTTIRIMTPIISGOSSEFMVVDLFTIPCDVYGPDKPVITWLLDDKPFTEGVVNE 1139  
QY 137 -G-----AH-----TI-A-N-----N----- 144  
Db 1140 DGSLLTIPVNEARGTFTCHAQNAAGNDRTVTLTHTPTTNAEQEKIALQNDIVLE 1199  
QY 145 -----HT--D-----MMB-----V-----D- 152  
Db 1200 CPAKALPPVPVRLWYEGEKIDSLPHITIREDCALVQNVKLENTGVFCVQSNLAGEDS 1259  
QY 153 -----G-DV-----EIP-----P-----NK-AV----- 163  
Db 1260 LSYLTWHEKPIISEVPGVDVWVGKFTIEIPCRATGVPEVIRTNKNGIDLKMDKKFS 1319  
QY 164 V-----LR-----E-----F-----E-----S----- 170  
Db 1320 VDNLGLTRIYEADKNDIGNVCVNTNEAGTSQMTTHVDVQEPPIILPSTQTNNTAVVGR 1379  
QY 171 -----E-----V-F-----E-----H-----E-----S----- 176  
Db 1380 VELKCYVEASPPASVTWFRGIAIGTDTKGYVVESDGLVIOQASVEDATIYTCASNEA 1439  
QY 176 -----A----- 176  
Db 1440 GKAEANLQVTVIASPDIKDPDVVTOBSIKESHPSLYCPVFNPLPQISWYLNKPLIDD 1499  
QY 177 -----W----- 178  
Db 1500 KTSWKTSDDKRLHVKAKITDGVYKCVARNAGSGSKFQVEVIVPLNLDSEKYYKV 1559  
QY 179 -----PVS----- 181  
Db 1560 FAKEGEVTLGCPVSGFPVQPNVWVDGTWBERGKYGKATLSNDGLTLHFDVSVKQEG 1619  
QY 182 -----D-L--LA----- 186  
Db 1620 NYHCAQSGNILDIDVELSVLAVPIVGEDNLEVLGKDISLSCDLQTESDDKTTFWMS 1679  
QY 187 -----GS-----DS-----T-----A-R----- 194  
Db 1680 INGSSEDRPDNVQIPSDGHRLYITDAKPNNGKYMCRVNTSAGAKARTLTDLVLEPPV 1739  
QY 195 -----IW--N-----LS----- 199  
Db 1740 EPVFEANQKLGNNPILQCVTGNPKPTVIWKIDGNDVDKSWLFDSELSLLRIEKLJTKG 1799  
QY 200 -----EN-----S-----T-----S-GS--T- 207  
Db 1800 SAQISCTAENKAGTASRDPFIQNIAPTFKNEGQDQETIIFRESEITILDCPVSLGDPQITW 1859  
QY 208 --Q-----L-VL--R-H--C----- 194  
Db 1860 MKQGLPLTENDALFTLDNLTRELILNARHDHDIYTCVANNTAGQVSKOPDVVVQVLPKIK 1919  
QY 217 -----EG-----G-----Q-DVP----- 223  
Db 1920 NAVVLEINEGEEIILTCDAEGNPTAKWDFNQGLDKPEAVFVNNHTVVNNVTKYHT 1979  
QY 224 -----SNK-----DV-----T-S-L--D----- 232  
Db 1980 GYVKYATNKVGQAVKTIINHVRTKPRFESGLTESELTVNLTRSIITLECDVDDAIGVGIS 2039  
QY 233 W-N-----SE-G-TL--L--A-T--GSY----- 245  
Db 2040 WTUNGKPLAETDGVQTLAGRPLHLIVSAKTDHDSIYACTVTNEAGVATKTNLFVQVPP 2099  
QY 246 -----DG----- 248  
Db 2100 TIVNEGGEYTVIENNSVLVPCVETGKPNPVWTKDGRPVGDLKSVQVLSGQFKIVHA 2159  
QY 249 -----A-----R----- 250  
Db 2160 EIAHKGSIYCMKNDVGTAEISFVDVDIITRPMIQIKINVIATKGAALPFKCPIDDDKN 2219

QY 251 -----IW-----T-----K-D-G-N----- 257  
Db 2220 FKQIILWLNYPIDLEAEDARITRLSNDRLTLILNTVTEDEQYSCRVKNDAGENSFDF 2279  
QY 258 -----L-----A-S-TL-----GO-H----- 265  
Db 2280 KATVLVPPTIIMLDKDKNKTAVEHSTVTILSCPATGKPEPDITWFKDGEAIHINIADIIP 2339  
QY 266 ----- 265  
Db 2340 NGELNGNQLKITRIKEDGAGKYTCADNSAGSVQDVQVNVNVTIPKIEKDGIPSDYESQ 2399  
QY 266 -----K-G-P----- 269  
Db 2400 NERVVISCVPVYARPPAKITWLGKAGPLQSDKFVKTSANGQLVLFKLRDSSKYTCIAT 2459  
QY 270 -----F-----A-----L-----K-----W----- 274  
Db 2460 NEAGTDKRDVKVSMVLPFSDPEFNIRITVNSGNPSTLHCPAKGSPSPITITWLKDGNAI 2519  
QY 275 --N-----K----- 276  
Db 2520 EPNDRYVPFDAGRQLQISKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPPIDGERREA 2579  
QY 277 -----K-G-N-----FI-----LSA----- 284  
Db 2580 VAVIEGFSSELCDSNSTGVVQKQDGLTINQDILRGDSFIQIPSSGKMKMFLSARKSD 2639  
QY 285 -----G-----V-----DK--T--TI-----I----- 292  
Db 2640 SGRYTCIVRNPAAGEARKLPDFAVNDPPSISDELSSANIQTIVPYVPVEINCVVSGSPHPK 2699  
QY 293 --W-----D-A-----H-----T-----G----- 298  
Db 2700 VYMLFDDKPLEPDSAAYELTNNGETLKIVRSOVEHAGTYTCEAQNNGVKARKDFLVRVTA 2759  
QY 299 -----E-----AK-----Q----- 302  
Db 2760 PPHFEKEREVAVRGDTMLLTCAESSVPLSSVYVWHAHDESQNGVITSKYAANEKTLN 2819  
QY 303 --Q-----F----- 304  
Db 2820 VTNIQLDDGEFYCTAVNEAGITKFKFLIVETPFLDQOKLYPIILGKRLTLDCSATG 2879  
QY 305 --P--F----- 306  
Db 2880 TPTPTILFMKDGKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGRSEKMMVEV 2939  
QY 307 -----H----- 308  
Db 2940 LLPPKLSKEWINVEVQAGDPLTLECPEDTSGVHITWSRQFGKDGQOLDMRAQSSDKSKL 2999  
QY 309 -----A-P-A-L-----D-----V-D----- 315  
Db 3000 YIMQATPEDADSYSCIAVNDAGBAVFOVTVPKIFGDSFSTTEIVADTTLEIPCR 3059  
QY 316 -----W-----Q----- 317  
Db 3060 EGIPPEISWFLDGKPILEMPGVITYKQGLSLRIDNIKPNQGRYTCVAENKAGRAEQDT 3119  
QY 318 -----S-----N-----N-----T-----F----- 322  
Db 3120 YVEISEPPRVWASEVMRVVEGRQTTIRCEVFGNPEPVVNLKDGEPYTSDDLQSTKLS 3179  
QY 323 -----A-----SC-----S--TD-----M--C- 330  
Db 3180 YLHLRETTLDGTVTCIATNKAGESQTTDVEVLVPPRIEDEERVLOQKEGNTVMVHCQ 3239  
QY 331 -----I-----HV-----C-----K 335  
Db 3240 VTGRPVVYVWKRNGKIEQFNPVLHIRNATRADEGKYSIASNEAGTAVADFLIDVFTK 3299

```
QY 336 -----L--CQ-----D-----RP-----I-K- 343
Db 3300 PTFETHETFNIVEGSAKIBCKIDGHPKPIISWLGGRPFNMNDNIILSPRGDTLMILKA 3359
QY 344 -----T-----F-----Q----- 346
Db 3360 QRFDGGLYTCVATNSYGDSEQDFKVVNVYTKPIDETIDQTPKAVAGGEIILKCPVLGNPT 3419
QY 347 -----G-----HT-----N---E---VN-----A-I- 355
Db 3420 PTVTWKGGDDAVPNDRSRHTIVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVIG 3479
QY 356 -----K----- 360
Db 3480 KPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFERGDKPVLYDRYSISPDGSHITI 3539
QY 361 -----G-----N-----L-----LA--S--C--S 368
Db 3540 NKAKLSGGKYICRASNEAGTSDIDLILKILVPPKIDKSNLIGNPLAIVARTIYLECPIIS 3599
QY 369 -----D--M----- 371
Db 3600 GIPOQDVITWTKNGMDINWTSRVILAQNNETFGIENVQVTDQGRYTCATNRGKASHDF 3659
QY 372 -----TL--K--I-----W---S---M-- 378
Db 3660 SLDLVSPPEFDIHGTQPTIKREGDTITLTCPIKLAEDIAQVMDVSWTKDSRALDGLDITD 3719
QY 379 -----K--Q----- 381
Db 3720 NVDISDPGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPPIVIDISRNDVQPOV 3779
QY 382 -----C-V-H-----D--L--O--O--H----- 390
Db 3780 AVNQPTIMRCVATGHPFPSIKWLKNGKEVTDENIRIVEGQVQLILRTSDHAGKWSCV 3839
QY 391 -----N-----KE-----I-Y-TI-----KW----- 399
Db 3840 AENDAGVKELEWLDVFTPPVWSKSDNPIKALGETITILPCNASGNPYQLKWKAGGSLI 3899
QY 400 -----SP-----T-G--P-----GT-----N----- 407
Db 3900 FDSFDGARIISLKGARLDIPHLKKTVDGYTCQALNAAGTSEASVSVDVLVPPPEINRDGID 3959
QY 408 -----N-----P----- 409
Db 3960 MSPRLPAQOQLTIQLCLAQKFPQMRWTLNGTALTHTSTFCITVASDSTFIQINNVSLSDK 4019
QY 410 -----N-A---NLN-----LA-----S--- 417
Db 4020 GVTYCAENVAGSDNLMLNVDVQAPFVINSNGGTQKQVIEGELAVIECLVEGYPAPQVSWLR 4079
QY 418 -----A-SFDS----- 422
Db 4080 NGNRVETGVQVRYVTDGRMLTIEARSLDSGIYLCASNEAGSAQOAVTLVLVSPKII 4139
QY 423 -----T-----VR-L-----W---DV----- 429
Db 4140 TSTPGVLTPSSGSKFSLPCAARGVPDPPIISWTINGNDIKDGENGHTIGADGTLHIEKAE 4199
QY 430 -----D-----R-----G-----I-C-I----- 435
Db 4200 RHLIYECTANDAGADTLFPVQTVIAPKISTSGNRYINGSEGTETVIKCEISESBFS 4259
QY 436 -----H-----T-L----- 440
Db 4260 WSKNGVPLPSSNLIIPSEYKLIKILSTRLSQGEYSCTAANKAGNATQTNLNVGVAPK 4319
QY 441 -----H--Q-----E-----P-V-Y-----SVAF--- 450
Db 4320 IMERPRTQVVKHQDQVTLWCEASGVQPAITWYKDNELLTNGVDETATTKKKSIFSSI 4379
QY 451 SP-----D----- 454
```

```
Db 4380 SPSQAGYTCYTKAENWVASTEEDIDLIVMIPPEVPPERMVNSTNPROTVFLSCNATGIPEP 4439
QY 455 -----R-----Y-----LA-----SGSF-----D-----K 464
Db 4440 VISWMEKSDNIAIONNEKYQILGTTILAIRNVLPDDDG-FPHCTIAKSDAGOKIATRKLIUNK 4498
QY 465 -----C-----VHI-----W-----N----- 470
Db 4499 PSDRPAPIWVECDCEKKGKPKKTEYMDRGDTDPDNPOLLPKWDESDSLNGSIAYRCMPGP 4558
QY 471 -----T-----Q-----T--CALV----- 477
Db 4559 RSRTVLLHAAPQFIVKPKNTTAAGIAVELRCSAAGPPHPTITWAKOGKLIEDSKFEIA 4618
QY 478 --H----- 479
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAVITCYE 4678
QY 480 -----Y-RG-T---G-----GI----- 486
Db 4679 RNQAVSRGLTWBYNGVPMFKNLGIFHMNNGSLVILDTSSLKEGDELYTCKVRNRHRS 4738
QY 487 -----PE-----V-C-----W----- 491
Db 4739 IPHLTSAFEGVEVKTIIDKVEVNGDSVVDCEVTSDDPLTTHVWTKNDQKMLDDDAIYV 4798
QY 492 -----N-----A-----A-----GD--K--V-- 498
Db 4799 LPNNSLVLNVKYGDEGVYKCVASNSIGKAPDPTQLNVVEGDFLPTGFEGSGINIDSS 4858
QY 499 ---G-----AS----- 502
Db 4859 NAGSSRREAYKENEDASTTTTTSPTTTTTPPLTTIIPALITLPAKQYPTDDYHE 4918
QY 503 -S---DG-----SV-----CV----- 509
Db 4919 GSANDGFGPTTQDSLFEFNPPLHPEISVNVTDCACTINENGDCVDKDGKTHNLKILIGE 4978
QY 510 -----LD----- 512
Db 4979 NHCPEGFAMNPHTRICEDLDECAFYQPCDFECINVDGFGQCNCPLGYELABEGCRDVNEC 5038
QY 513 ---R---K 514
Db 5039 ESVRCEDCK 5047

RESULT 3
I38346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jul-2004
C:Accession: I38346
R:Labelit, S.; Kolmerer, B.
Science 270, 293-296, 1995
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUD:96026330; PMID:7569978
A:Accession: I38346
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:CROSS-references: UNIPROT:Q10465; EMBL:X90569; NID:gl017426; PIDN:CAA62189.1; PID:gl017426;
C:Genetics:
A:Gene: GDB:TTN
A:CROSS-references: GDB:127867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 70.9%; Score 2594.7; DB 2; Length 7962;
Best Local Similarity 8.7%; Pred. No. 1.8e-52;
Matches 450; Conservative 52; Mismatches 9; Indels 4663; Gaps 382;
QY 1 MS-----IS-----V-N----- 9
```



QY 215 -----IRE----- 217  
Db 2276 RFVSKNSLTVVAGEPAELQASIEGAQPIFVQWMLKEKEVIRESENIRITFVENVATLQF 2335  
QY 218 -----GG----- 219  
Db 2336 AKAEFANAGKYICQIKNDGGMRENMATLWLEPAVIVEKAGPMITVVGECTICLECKVAGT 2395  
QY 220 -----QD----- 221  
Db 2396 PELSVYKDGKLLTSSQKHKFSYINKISSRLILSVEROQAGTYTTFQOVNNVGKSSCTAV 2455  
QY 222 -----V-PS-----N-----K----- 226  
Db 2456 VDVSRAVPSPFTRRLKNTGGVLGASCILECKVAGSPISVAMFHEKTKIVSGAKYQTTF 2515  
QY 227 -D-V-T-----SLD----- 232  
Db 2516 SDNVCTLQLNSLSDSDMGNYTCAANVAGSDECRVLTVOEPPSFVKEPEPEPLEVLPCKNV 2575  
QY 233 -----W-----N-----S-E----- 236  
Db 2576 TPTSVIRGTPFPKYNWFRGARELVKGRDRCNIYFEDTVAELEFNIDISQSGEYTCVVSN 2635  
QY 237 -G-----T-LL-----A-----TG-----S-Y--DGF----- 248  
Db 2636 AQOASCTTRLFVKEPAALFKRLSDHSVEPKSIILESTYTGTLPISVTWKOGFNITTS 2695  
QY 249 -----A-R----- 250  
Db 2696 KCVITTEKTCILEILNSTKRDAGQYSCIEAGRDVCGALVSTLEPPYFVTELEPLEA 2755  
QY 251 -----I-W-----TK-----D-G----- 256  
Db 2756 AVGDSVSLQCVAGTPEITVSWYKGTDLRPTPEYRTFTNNVATLVFNKNVINDSGEY 2815  
QY 257 -----N-L--AS-----TL-----G-----Q----- 264  
Db 2816 CKAENSIGTASSKTVFRIQERQLPPSPARQKDIEQTVGLPVLTCRLNGSAFIQVCWYR 2875  
QY 265 -----H-----K-----G----- 267  
Db 2876 DGVLLRDHENLQTSFVDNVATLKILQTLDSHQYSCSASNPLGTASSSARLTAREPKKS 2935  
QY 268 -----P-----I-----F-----A--LK--W--NK----- 276  
Db 2936 PFFDIKPVSIDVIAGESADFECHVTGAQPMRITWSKDNKEIRPGNYTITCVGNTPHLRI 2995  
QY 277 -----KG----- 278  
Db 2996 LKVGKDGSGQYTCQATNDVCKMCSAQLSVKBPFPVKLEASKVAKQGESIOLECKISG 3055  
QY 279 -----N-FI-----L-----SA-----GV-D----- 287  
Db 3056 SPEIKSVWRNDSSELHESWKYNNSPFINSVALLTINEASAEKSDGYICEAHNGVGDASCST 3115  
QY 288 -----K-----T-----T-----IIW----- 293  
Db 3116 ALTVKAPPVFTQKPSVPGALKGSDVILQCEISGTPPPFVWVWVKDKQVRNKKFKITSKH 3175  
QY 294 -D-----A-----H----- 299  
Db 3176 FDTNLHLNLEASDVGEYHCKATNEVGS DTCSCSVKPEPPFPVKKLSDTSTLIGDAVEL 3235  
QY 300 -A-----K-----Q----- 303  
Db 3236 RAIVEGFQIPISVWVLKDRGEVIRESENTRISFIDNIATLQLGSPEASNSGYICQIKNDA 3295  
QY 304 -----F-----P----- 306  
Db 3296 GMRCSAVLTLEPARIIEKPEMTVTGPNPFALCECVTGTPELSAKWFKDGRLELSADSK 3355

QY 307 -H-S-----A-----P-A-----L-----D-V----- 315  
Db 3356 HHITFINKVASLKI PCAEMSDKGLSYFEVKNVSGKNCCTVSVHVSDRIIVPPSFIRKLKDV 3415  
QY 316 -----W-----QS-----N----- 319  
Db 3416 NAILGASVVLCEKRVSSAPISVCGWFOGNEIVSGPKQCSQSPSENVCTLNLSLLEPSDTGI 3475  
QY 320 -----N-----TF----- 322  
Db 3476 YTCVAANVAGSDECSAVLTVOEPPSPETQSDSVLELPGMSLTFTSVIRGTTPPFVKWFKG 3535  
QY 323 -----ASC----- 325  
Db 3536 SRELVPGESCNISLEDFVTELELFEVQPLESGDYSCLVNDAGSACTTHL FVKEPATFV 3595  
QY 326 -----S-----T-----D-----M----- 329  
Db 3596 KRLADFSVETGSPVLEATYTGTPPISVSWIKDEYLIQSERCISITWTEKSTILEILEST 3655  
QY 330 -----C-----I-----HV-----CKL-G----- 337  
Db 3656 IEDYAOYSLIENEAGQDICEALVSVLEPPYFIEPLEHVEAVIGEPA TLQCKVDGTPEIR 3715  
QY 338 -----Q-----D----- 339  
Db 3716 ISWYKHTKLRSA PAYKQPKNVASLVINKVDHSDVGEYSCKADNSGVAVASSAVLVIK 3775  
QY 340 -R--P-----I-KT-----F-----Q-----G-----H-- 348  
Db 3776 ARKLPFPFARKLVHETLGFVPVAFECRINGSEPLQVSWYKGVLLKDDANLQTSFVHN 3835  
QY 349 -----T-----N-----EV----- 352  
Db 3836 ATLQILQTDQSHIGQYNCASNPLGTASSAKLILSEHEVPPFDLPKPSVDLALGESGT 3895  
QY 353 -----NA-IK--W--D-----PTGN-----L--L-----A--SC--S-- 368  
Db 3896 FKCHVTGTAPIKITWAKONREIRPGNYKMTLVENTATLTVLKVGKGDAGQYTCYASNTA 3955  
QY 369 --D-----DM--T-L--KI-----W-----SMK-Q-- 380  
Db 3956 GKDSCSAQLGVQBPFRFIKLEPSRIVKQDEFTRECKIGSGPEIKVLWKDET-EIQS 4014  
QY 381 -----D-----N-----C-----V-----H-- 385  
Db 4015 SKFRMSFVDSVAVLEMHNLVSEDSGYTCEAHNAAGSASSSTSLKVKEPPIFRKKPHPIE 4074  
QY 386 -----D-----LQ-----Q-----H----- 390  
Db 4075 TLKGADVHLECELOGTTPPHVSWYKDKRELRSKKYKIMSENFLTSIHLNVDAADIGEY 4134  
QY 391 -----N-----KEI-Y--TI-----K 398  
Db 4135 QCKATNDVGS DTCVGSIALKAPPRFVKKLSDISTVVGKEVQLQTTIEGAEPI SVVWFKDK 4194  
QY 399 -----W-----S----- 400  
Db 4195 GBIVRESNIIWISYSENATLQFSRVEPANAGKYTCQIKNDAGMOECFATLSVLEPATIV 4254  
QY 401 --P-----TG-----P-----G--T-----N-----N-- 408  
Db 4255 EKPESTKVTTGDTCTLECTVAGTPELSTKWFKDKGELTSDNKYKISFFNKVSGLKLIINVA 4314  
QY 409 P-----N-----A-----N-L-----M-----L-- 415  
Db 4315 PSDSGVYSEVQNPVGKDSCTASLQVSDRTPPSPFRKLKETNGLSGSSVWMECKYVGS 4374  
QY 416 -ASAS-F-----DS----- 422  
Db 4375 PISVWFHEGNEISSGRKYQTTLTONTCA TVNMLEESDGDYTCIATNMAGSDECSAPL 4434  
QY 423 TVR-----L-----W----- 427





```
QY 1 M-----SISD-----EV--N-----F-----L-----11
Db 20024 MRNTEITAENLTINUKESVTADAGYEYETAANSSGTTKAFINIVVLDPRPGPTGPPVIS 20083
QY 12 -----Y-----R--Y--L--Q--S-----G-----20
Db 20084 DITEESVTLKWEPPKPDGGSQVNTYILLKRETSVAVTEVSATVARTMKVWKLTTGEEY 20143
QY 21 -----F-----S-----H-----SA 25
Db 20144 QFRIKAENRFGISDHIDSACVTVKLPYTPPGPSTPWNTVNTRESITVGMHEPVSNGGSA 20203
QY 26 -----F--T-----F-----G--K--SH-----33
Db 20204 VVGYHLEMDRNSILWOKANKLVIRTHFKVTTISAGLIYFRVYAENAAGVKPSHPSE 20263
QY 34 -----F-----F-----G-----33
Db 20264 PVLADACEPPRNVRITDISKNSVLSWQAPAFDGGSKITGYIVERDDLPGRWTKASFT 20323
QY 34 -----IS--Q--S-----N-----I--N-----G-----41
Db 20324 NVTETOFTISGLTQNSQYFRFARNAVGSIENPSEWVGPIITCIDSYGGEVIDLPLEYTE 20383
QY 42 -----A-----L--V--P--P-----A-----A--LJ-----50
Db 20384 VVKYRAGTSVKLRAGISGPAPTIEWYKDDKELQTNALVCVENTTDLASILKADRLNS 20443
QY 51 -----S--I--Q--K-----G-----L--Q-----58
Db 20444 GCVELKLRNMAASATIRVQILDKPGPGPIETKVTVAEKITLLWRPPADDDGGAKITH 20503
QY 59 Y-VEA-E-----VS-----I--NE-----D-----69
Db 20504 YIVERKTSRVMSVMSEHLEBECIITTKIKGNEYIFRVAVNKYIGIEPLESDSVAK 20563
QY 70 -----G-----T-----L--F-----73
Db 20564 NAFVTPGPGIPEVTKITKNSMTVWSRPIADGGSDISGYFLEKRDKKSGLWFKVLKETI 20623
QY 74 D-----G-----R-----PI-E-S-LSL-----ID-----A-----86
Db 20624 RDTRQVKTGLTENS DYQYRVCAVNAAGQGFSEPSF-YKADPIDPPGPPAKIRIADST 20682
QY 87 -----V-----MP--D-----VW--Q-----T-----94
Db 20683 KGSITLWSPVYDGGSAVTGYVVEIROGEEBETWTVTKGEVTTTEYVYVSNLKPQVNY 20742
QY 95 -R-----Q-----QA-----OA-----98
Db 20743 FRUSAVNCAGQGEPIEMNEPVOAKDILEAPETDLVALRTSVAKAGEDVQVLIIPKGRP 20802
QY 99 -----YR-D-K-L--A-----Q-----105
Db 20803 PPTVTRKDEKNLGDARYSIENTDSSLLTIPQVTRNDTGKYLFTIENGVGPKSSTVS 20862
QY 106 -----Q--QA-----A-----AA-----110
Db 20863 VKVLDTPAACQKLQVXVSRGVTLLWDPPLIDGGSPIINYVIERKDATKRTWSVVVSHKC 20922
QY 111 -----A-----A-----AA-----A-----114
Db 20923 SSTSFKLIDLSEKTPFFFRVLAENEIGICEPCETTEPVKAAEVPAPIRDLMSKDKSTKTSV 20982
QY 115 -----AA--A-----AA--Q-----121
Db 20983 ILSWTKPFDGGSVITEYVVERKGEGEQTWSHAGISKTCIEVSQLKEQSLEFRVFAKN 21042
QY 122 --G-S-----AK-----N-----126
Db 21043 EKGLSDPVTIGPITVAKELIITPEVDLSIDIPGAQVTVRIGHNVHLELPYKGPXPSWLK 21102
QY 127 -G-----EN--T-----A-----N-----132
Db 21103 DGLPLKESEFVRFSTENKILTSIRNAKKEHGGKYTVILDNAVCRIVAPITVITLGPSPK 21162
QY 133 --G-----E-----E--NG-----AHT-----IA-----142
Db 21163 PKGPFRFDEIKADSVILSWDVPEDNGGGEITCYSIEKRETSQTNMKWCVSSVARTTFKVP 21222
QY 143 N-----N-----H-----T-D-M-----148
Db 21223 NLVKDAEQFRVRAENRYGVSQLPVSSIIIVAKHQFRIIPGPKPVIIYNTSDGMSLTWDA 21282
QY 149 -----ME-----150
Db 21283 PVDGSGSEVTGFHVEKKERNILWQKNTSPISGREYRATGLVEGLDYOFRVYAENSAGL 21342
QY 151 -----VD--G-----DV--E--I-----PP-----NK--161
Db 21343 SSPSPSKFTLAVSPVDPPTDYIDVTRETITLWNPPLRDGGSKIVGYSIEKQGNR 21402
QY 162 -----AV-----V--164
Db 21403 WVRCNFTDVSCQYTVTGLSPGDRIYEFRIIARNVAGTISPPSQSSGIINTRDENVPIVE 21462
QY 165 -----LR-----G-----167
Db 21463 FGPEYFDGLIIKSGESLRIKALVQGRPVPRVTWPKDGVIEKRMNWEITNVLGSLTVR 21522
QY 168 -----H-----E-----S--E-----V--F--I-----174
Db 21523 DATDRHGVYTVAKNAGSAGAKAEIKVKVQDTPGKVGVPFRFTNITGERKMTLWMDAPLND 21582
QY 175 -C-----AW-----N-----178
Db 21583 GCAPITHYIIEKRETSRLAWALIEDKCAQSVTAIKLINGEYQFRVSANVKGVRPLD 21642
QY 179 --PV-----SD-----L-----183
Db 21643 SDPVVAQIQTVPDAPGPEPSNITGNSITLTWARPESDGGSEIQOYILERREKSTRWV 21702
QY 184 -----L-----ASG--S-----G-----189
Db 21703 KVISKRPISERFKVTGLTEGNEYEFHVMAENAGVPASGISRLIKCREPVNPGPTV 21762
QY 190 -----D--S--T-----AR-----I-----W--N-----LS-----199
Db 21763 VKVTDTSKTTVLEWSKPVFDGGMELIIGYIEMCKTDLGWHKVAEACVKTRYVTDLQ 21822
QY 200 --E-----N-----S--TSGS-----T-----Q-----L--209
Db 21823 AGEYKFRVSAINGAGKGDSCVET--GTIKAVDRLTAPELIDIDANFKQTHVVRAGASIRLF 21881
QY 210 -----V-----LR--H-----C-----214
Db 21882 IAYQGRPTPTAVMSKPDNSLRLADIHTDTSFSTLTVCNCRNDAGKYTLTVENNSGSKS 21941
QY 215 -----I-----R-----EGG-----219
Db 21942 ITFTVKVLDTPGPGPIITFKDVTGRSATLMDAPLDDGGARIHHYVVEKREASRSQVVI 22001
QY 220 -----Q-----D--VP-----S--NK-----DV--TS 230
Db 22002 SEKCTRQIFKNDLAEGVPPYFRVSAVNEYGVGPEVEMPEIVATEQAPPRRLDVEDTS 22061
QY 231 -----L-----D-W-----233
Db 22062 KSSAVLWMLKPDHGGSRITGYLLEMRQKGSDLWVAGHTKOLTFTVERLVEKTEYEFV 22121
QY 234 --N-----SE-----G--T--LJ--A-----T-----G-----243
Db 22122 KAKNDAGISEPREAFSSVIIKEPQIEPTADLTGINTQLITCKAGSPFTIDVPSIGRPAPK 22181
QY 244 -SY-----D-----G--F-----A-----249
```

Db	22182	VTKLEEMRLKTRDVSITTTKORTTLTVKDSMRGDSGRFYFLTENTAGVKTFPSVTVVVI	22241
Qy	250	-R-----I-----W--TKOG-----NL-----ASTLG-----	263
Db	22242	GRPGVTPGPIEVSSVSAESCVLSWGBPKDGGGTEITNYIVEKRES--GITAMQLVNSSVK	22299
Qy	264	-Q-----H--K-----G-----PI-----F-----A	271
Db	22300	RTQIKVTHLTXYMEYFSRVSSENRFGVSKPLESAPIABHPFVPPSAPTREPVYHVSANA	22359
Qy	272	-L-----K-----W--NK-----	276
Db	22360	MSIRWBEFYHGGSKIIGYWEKERNITLWKENKVPCLCNYKVTGLVEGLEYOFRY	22419
Qy	277	-----G-----	276
Db	22420	ALNAGVSKASEASRPIMAQNPVDAQGRPEVTDVTRSTVLSIWSAPAYDGGSKVGYIIE	22479
Qy	277	-K-----G-----NF--I--LS-----AGV-----	286
Db	22480	RKPVSEVGDGRWLKCNVTIVSDNFFVTALSEGDTVEFRVLAKNAAGVSKGSESTGPVT	22539
Qy	287	-----DK-----	288
Db	22540	CRDEYAPPKAELDARLHGLDVLVTRAGSDVLVDAVGGKPEPKIITWTKGDKELDCEKVL	22599
Qy	289	-T-----T-II-----	292
Db	22600	QYTKRATAVIFKCDRSDSGKYTLTVKNASGTVKAVSMVWVLDSPGCGKLTVSRVTQEK	22659
Qy	293	-W-----D--A--H-----T-----GE-----	299
Db	22660	CTLAWSLPQEDGAEITHYIVERRET SRLNMWIVEGECPTLSVYVTRLIKNEYIFRVEA	22719
Qy	300	-----A-----K-----Q-----	302
Db	22720	VNKYGPVPESEPIVARNSTIPSPGPEVGTGKEHIIIQWTKPESDGGNEISNVLV	22779
Qy	303	-----OF-----P-----F-----	306
Db	22780	DKEKESLRWTRVKNKYVVDYDTRBLKVTLSMEGCDYQFRVTANNAAGNSEPERSNFISCR	22839
Qy	307	-----HS--A--PALD-----V-----D-W-Q--SN--	319
Db	22840	EPSYTPGPPSAPRVDTTKHSISLAWTKPMYDGGTDIGVYVLEMQEKOTDQWVRVHTNAT	22899
Qy	320	-NT-----F-----	322
Db	22900	IRNTEFTVPLKMGQKYSFRVAANVKGMSSEYSEIAEIPVERIBIPDLELADLKKTV	22959
Qy	323	-----AS-----	324
Db	22960	TIRAGASLRMLVSGRPPPVITWMSQGDILASRAIIDTSTESYLLIVDKVNRDAGKYT	23019
Qy	325	-----C-----S-----T-D-----	328
Db	23020	IEAQSGKKSATVLVKVYDTPGPCPSVKVKEVSRDSVTITWETPTIDGAPINNYIVEK	23079
Qy	329	-M-----C-----I-----H-----	332
Db	23080	REAAAFKVTTKCSKTLRYISGLVEGTMHYPVLPENIYIGIEPCETSDAVLVSEVPL	23139
Qy	333	-----V--CK-----	335
Db	23140	VPAKLEVVDVTKSVTLWAEKPLYDGGSLRTGVYLBACKAGTERMMKVVTLPKPTVLEHTV	23199
Qy	336	-----L-----G-----QD--R--P-I-----KT-----	344
Db	23200	TSLNEGOYLFRIPRAQNEKGVSEPRETVTAVTVQDLRVLPITDLSTMPQKTHVPAGRPV	23259
Qy	345	-----FOG-----HT-----	350

QY 461 SF-----D-----K-----C-V-H- 467  
 Db 24399 EFRVCENLGGSEWSEIPIPKSDVPIQAPHFKEELRNINRYOSNATLVCKVTGHP 24458  
 QY 468 -I-W-----NTQ-----TG- 474  
 Db 24459 KPIVKYRQKEIIADGLKYRIQRFKGGYHQLIIASVTTDDATVYQVRATN-QGGSVSGT 24517  
 QY 475 -----A-----L-----VHSYRG-----T-----G----- 485  
 Db 24518 ASLEVEVPAKIHLPKTLKLEGMGAHVALGEVVSIIKIPFSGKPDVITWQKQODLIDNNHY 24577  
 QY 486 -I-----F-----E-----VC----- 490  
 Db 24578 QVIVTRSTSLVFPNGVERKDAGFYVCAKMRFGIDOKTVELVDVADPPRGRVKVSDAS 24637  
 QY 491 -----W-----N-----A----- 493  
 Db 24638 RDSVNLWTPEPASDGGSKIITNIVEKCATTAERMLRVQARETRYVINLFGKTSYQPRV 24697  
 QY 494 -A-----G-----DK-----V----- 498  
 Db 24698 IAKNFKLSKPSSEPTITKEDTRAMNYDEEVDREVSMTKASHSTKELYKYMIA 24757  
 QY 499 -----G-----AS-----A-----SD----- 504  
 Db 24758 EDLGRGFGIVHRCVETSSKTYMAKFKVKGTDQVLVKKEISILNARHNILHLHESP 24817  
 QY 505 -----G-----S-----VC-----V-----LDLR- 513  
 Db 24818 ESMEELVMIFISGLDIFERINTSAFELNREIVSVYHVQCEALQFLHSHNIGHFDIRP 24877  
 QY 514 -----K 514  
 Db 24878 ENIIYQTRRSSTIK 24891

RESULT 5  
 T30192  
 Probable peptide synthetase - Aureobasidium pullulans  
 C:Species: Aureobasidium pullulans  
 C:Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 09-Jul-2004  
 C:Accession: T30192  
 R:Peery, R.B.; Thornewell, S.J.; Tobin, M.B.; Skatrud, P.L.  
 submitted to the EMBL Data Library, January 1997  
 A:Description: Discovery of an MDR-like gene adjacent to a peptide synthetase in Aureoba  
 A:Reference number: Z20767  
 A:Accession: T30192  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-10797 <PEE>  
 A:Cross-references: UNIPROT:O94116; EMBL:U85909; NID:g4099310; PID:g4099313; PIDN:AAD005  
 C:Genetics:  
 A:Introns: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2  
 C:Keywords: carrier protein  
 F:1618-1688/Domain: acyl carrier protein homology <ACP1>  
 F:3682-3752/Domain: acyl carrier protein homology <ACP2>  
 F:5615-5685/Domain: acyl carrier protein homology <ACP3>  
 F:7503-7573/Domain: acyl carrier protein homology <ACP4>  
 F:9683-9752/Domain: acyl carrier protein homology <ACP5>

Query Match 70.6%; Score 2583.7; DB 2; Length 10797;  
 Best Local Similarity 8.4%; Pred. No. 6.5e-52;  
 Matches 459; Conservative 41; Mismatches 9; Indels 4923; Gaps 395;  
 QY 1 M-----SI--S-----D----- 6  
 Db 4302 MLNTQVIOQFGSGISGILYMGVPTAAIHCTLPQFGVDLPAGTIGIPLDVTSCFIV 4361  
 QY 7 -----E-V-----N-----FLV-----YR--- 14  
 Db 4362 KPTESTKHASQLEILPGEIGELVIGGHQADGLYNREBQTRAAF-VTHPKFGGLYRTGD 4420

QY 15 -----Y-----L-Q-----E-----SG-----F-- 21  
 Db 4421 KARLHRNTGLECYGRISSGQVQLRGQVELGEIEHAASKAGGCHAVIASVISGLLVLCI 4480  
 QY 22 --SH-----SA-----F-----T-F----- 28  
 Db 4481 GDHRVSSKDIKACQKWLPAYMIPSDIVLLDDFPYLPBGKVDKKLETDYNSNTAQHVS 4540  
 QY 29 G-----IKS-----H-----I-----SQ----- 36  
 Db 4541 GSSDLSENAREITRIIESVLGVSIDHSTDLASAAGLSLRAIOVASQLRRQGCADLGALEL 4600  
 QY 37 -----SNIN-----G----- 41  
 Db 4601 LSVSNVLADELVRKADESININDNSEKWKQTVHELRSVERDFESKAFVSGIEDVLP 4660  
 QY 42 -----A-LV-----P-P-----A-----AL-----IS- 51  
 Db 4661 TPLQDAMLVETAKRPOAYCNELELTVPKIPVERVQALFALAORHTALRSRGMPSGVSH 4720  
 QY 52 -----II-----Q-K-----G-----L-QY-----VEAE--V----- 64  
 Db 4721 CAYTOVWKTLVTSQFAHVKSFTTGWSTNRETLLRPLHFQYKCSGAEREILVAIHALLY 4780  
 QY 65 --S--I-----NE-----D--G- 70  
 Db 4781 DQWSEVILEDLTLNERTPERPSGAVNKFNLRRSEDOTSHLDFWGEYLSVDTVPCR 4840  
 QY 71 -----TL-----F-----D- 74  
 Db 4841 LPNLSPKMPPQPLQSIQHTIEMMETLQAASHYSCSAHFVFAAYAILLGYMGTEDT 4900  
 QY 75 -----GR--P--IES-----LS-L--I-----DA-----VM----- 88  
 Db 4901 VFGTFSGRTLPIVETESVMGPLLSTLPTPRINTLESRKFSVLSRLQEDNRKIMRHSMTS 4960  
 QY 89 -----P-----D-V-V-O-T-----R-----Q- 97  
 Db 4961 LADIKACGFNPGAEVDFSIFFWQETARPDARAQTLNLNVEAHYLEFNLTELEPTQOG 5020  
 QY 98 -----A-Y-----R-----DKL-----A-----Q- 105  
 Db 5021 VKTKATYQSSLLPLQHVKTLLQQLDALVKIIVARPETHMNEISDQLPISVLSVANSEPOS 5080  
 QY 106 -----Q-----Q-----Q- 107  
 Db 5081 FVYKAGLSLVENHALNNSGGLALVPAHDIREGTSRMESLTYGELNTRANQLANLISQG 5140  
 QY 108 A-----AA-----AA-----A-----A----- 113  
 Db 5141 AKRDELICVCMKESVSLYLILAAVKAGCGYLPVPETPAAIRQILAEADVFKCLTDSS 5200  
 QY 114 -A--A-----A-----A-----A----- 118  
 Db 5201 MAPVIADVSRCHIMNVDTTDCSAQSCGQPLDFKPTDIAVAVFTSTGTPKPGVLVTQEN 5260  
 QY 119 -----S--Q--Q-----Q-----G-----SA-----K- 125  
 Db 5261 ILSNLEVLKSIYPVPEGSRLQACNQAFDVSVEIFFTWYTGMLCSASKDVMFRDFEKA 5320  
 QY 126 -N-----GE----- 128  
 Db 5321 INELETHLSLTPTVAULTDPAHPRVKFLVTAGEAVTHVHVHGAWAGKLYOYGPSETT 5380  
 QY 129 N-T--A-----N-GE--EN-----GA----- 138  
 Db 5381 NICTVNSAVESDHVINNIGPAPENTSFAVLTOGDDFQLVPLGGLGELCGGQQVFRGYQN 5440  
 QY 139 -----H-----TI----- 141  
 Db 5441 MPELTSKIINPNYGRIVRSGLGRLLPDGTLLIQGRTDDQKTRGRIEUGEISGCULL 5500

QY 142 ---AN---N---H--- 145  
DB 5501 QPSPVQNCNCAIEVKTADKERLMAFIPSGVSKDSYSILOPKNLEBIIKSIYAHLANLP 5560  
QY 146 ---T---DM--- 148  
DB 5561 AYMPDALVPVSAIPOTSQKIDKRRLASGALSALTVEDLNAYSRGADDDTSELSATEQO 5620  
QY 149 ---M-E-V---D---G---DV-EI---P--- 158  
DB 5621 LASALADTLQMSQTSIGRSTSPALGLDSVSAIRLATNLRLKEYGYSIDVSQILKRTIAR 5680  
QY 159 ---P---NK---AV---VL--- 165  
DB 5681 LAPLLGESSKOTNEPVTADCEAAVGSYLHDESWSQLHEHGTVSQVLPCTPLQEAMLS 5740  
QY 166 -R-G---H---E---S-E-V-PI--- 174  
DB 5741 ARDTGSSAYRNKTLFSLHGSVDKLCACWEVMLQRHDIILRTIFLSTEDSRFPFVQAVLSQ 5800  
QY 175 ---C---A---W--- 177  
DB 5801 WTLPWQECDDIPDLSTLLDSAKAGGDSIVDHSPPMKIQVYRSESTVYLLDMHHALEYDA 5860  
QY 178 ---N---PVS--- 182  
DB 5861 NAMSNNLYEVEQLYKQSLAPVSPKPFNFMISTSVBEADALFRDQLREFVPKPKRTD 5920  
QY 183 ---L---LA--- 187  
DB 5921 VKSGFTITGRNLNYSKPMVETFLSKHSTTMLSLTQAMWMTLAASQSYSVDCGNGVSGR 5980  
QY 188 --- 187  
DB 5981 SVPVDGIESIVAPCFNTIPVRVDLSKHSNGLVVKALQRWNIDSLPYQLTPLRQAOAG 6040  
QY 188 ---SG-D---ST--- 192  
DB 6041 TNGKRLFDLSVLVLLQDDTDLDSAINRLEGESGMDMYTSGTGLPKAVCLSHRAVTSQL 6100  
QY 193 A---R---I-W---N--- 197  
DB 6101 AHDRTFIPSRFLQFASPTFDVSVEIFPPWYRGATLVSVERNRLGLDLPGTITSINIDA 6160  
QY 198 ---L---S---EN---S---T---SG--- 205  
DB 6161 ABELTSVAASLVHRHENVTLRALLTIGEMLNTQVIOQFGSGSGILYGMVGPTEAAI 6220  
QY 206 ---ST--- 209  
DB 6221 HCTLPQFGVDLPAGTIGIPDVTSCFIVKPTSTKHAQLEILPIGEIGELVIGGHOLA 6280  
QY 210 ---V---L---R-H---C-I---R--- 216  
DB 6281 DGYLANREEQTRAAVTHPKFGLGYRTGDKARLHRNGTLECYGRISSGQVKLRQORVELGE 6340  
QY 217 -E---GG---O---D-V--- 222  
DB 6341 IEHAKAGGCHAVIASVISGLLVLEFCIDGPHRVSSKDIKSAQKWLPAWIPSDIULLD 6400  
QY 223 ---P---SN---KD---V---T--- 229  
DB 6401 DFPYLPSPGVKKRLETDNSNTAQHVSGSDLSSENAREITRIESVLGSDHSDTDLA 6460  
QY 230 ---SL---D---W--- 233  
DB 6461 AGLDLSLRAIQVASQLRROGCADLGALELLSVSNVLALDELVRAKADESNINDNSEKWKQ 6520  
QY 234 ---N---S--- 235  
DB 6521 TVHELRASSVERDFESKAFVSGIEDVLPTPLQDAMLVETAKRPOAVCNELRVTSPKIPV 6580  
QY 236 E---G---TL---LA---TG--- 243

DB 6581 ERVROALFALQRTALRGFMPSGVSHCAYTQVINKTLVTSQFAHVKSFTTGWSVTNRE 6640  
QY 244 ---S---YD---G--- 247  
DB 6641 TLLRPLHFOYKCSGABEILVAIHHAHYDQMSVEVILEDETLLOQNERTPERSFGAVNK 6700  
QY 248 -FA-R---I-W---TKD---G-NL--- 258  
DB 6701 FPNLRSEDOTSHLDPWGEYLS-DVTPGRLPNLSPKMPMPQPLQSIQHTIEMDMETLROA 6759  
QY 259 ---A---S-TL---G--- 263  
DB 6760 AHSYCSAHVFFQAAVAILLGFYMGTEDTVFGTVFSGRTLPIVEIBSMVGLLSTLPTRI 6819  
QY 264 ---Q---H---K-G-P---IPALKW--- 274  
DB 6820 NTLERKSPDVLRLQEDNRKMRHSMSTLADIKKACGNPGEAVPDSIFV-WQETARP 6877  
QY 275 ---N---K--- 276  
DB 6878 DARAQTLLNLVEAHDYLEFNLTLELEPTQGVKTKATYQSSLLPLQHVKTLLQLDALVK 6937  
QY 277 ---K-G---N---F--- 280  
DB 6938 IVVARPETHNEISDQLPISVLSVANSEPSQSFYKAGLGSVENHALNNSGGIALVFAHD 6997  
QY 281 ---I---LS---AG-285  
DB 6998 IREGSRMESLTYGELNTRANQANLYISQGAQRDELIQVMEKSVLSLSILAAKAGC 7057  
QY 286 ---VD--- 287  
DB 7058 GYLPLVPETPAARIRQILAEADVKECLTDSSMAPVIADVSRCHIMNVDTDCSAQSGTGP 7117  
QY 288 ---K---TT---I--- 291  
DB 7118 QLDFKPTDIAYAVFTSGTTGPKGVLVTQENILSNLEVLISKIYPVPEGSRLQACNOAFD 7177  
QY 292 ---I-W---D-AH--- 296  
DB 7178 VSVFIFFTWYTMCLCSASKDVMFRDFEKAINELEITHLSLTPTVAALTDPAHIPRVKF 7237  
QY 297 ---T-GEA---K-Q--- 302  
DB 7238 LVTAGEAVTHVHGAWAGKLYQGYCPSETTNICTVNSAVESDHVINNIGPAPENTSAPV 7297  
QY 303 ---Q---F---P---F---H---SA---P 310  
DB 7298 LTQGDQDFQLVPLGLGELCFGGQVFRGYQNMPELTESKIIINHNYGRIYRSGDLGRLLP 7357  
QY 311 ---AL---D---V---D---W--- 316  
DB 7358 DGTILIQRTDDORKIRGORIELGBISGCLLOFPSPVQNCNCAIEVKTADKERLMAFIPSG 7417  
QY 317 ---Q---S---N--- 319  
DB 7418 YSKDSYSILOPKNLEEEIKSIYAHLANLPAYMVPDALVPVSAIPQTSQKIDKERLAS 7477  
QY 320 ---N---T-FA--- 323  
DB 7478 DGSALTVEDLNAYRGADDDTSELSATEQQLASALADTLQMSQTSIGRSTSPFALGLDS 7537  
QY 324 --- 323  
DB 7538 VSAIRLATNLRLKEYGYSIDVSQILKRTIARLAPLLGESSKOTNEPVTADCEAAVGSYL 7597  
QY 324 ---S-C---S---T---DM---C-330  
DB 7598 HDESUVSQLHEHGQTVSQVLPCTPLQEAMLSARDTSGSSAYRNKTLFSLHGSVDKLCACW 7657  
QY 331 -I---H---V---C---K---L---G--- 337

Db 7658 EVMLQRHDIILRTIFLSTEDSRFPFVQAVLQWTLPMQECDDIPDQLSTLLDSAKAGDSI 7717  
QY 338 -----T-----Q-----DR---PI--K--- 343  
Db 7718 VDHSPPMKIQVRSSESTVYLLDMHHALYDANAMSNLLYEVEQLYKQDSLSAPVSPFL 7777  
QY 344 -----T-----F--Q-----G-----H--- 348  
Db 7778 NFMISTVEADALFRDLREFVPPKPTDVKSGFTITGRNLNYSKMWETFLSKHSTT 7837  
QY 349 -----T-----N-----E-----V----- 352  
Db 7838 MSLTQAMMKTLAASQSDVCCGVNVSGRSVFVPGIESLVAPCFNTIPVRVDSLKHRS 7897  
QY 353 N-----A-----I-----K-----W----- 357  
Db 7898 NLGLVKALQRWNIDSLPYQLTPLRRIQAQAGTNGKRLFDSLVLLOQDTTDLDSAIRLEG 7957  
QY 358 -----D-----PT-----GNL---L---ASC-----SD 369  
Db 7958 ESGVMDPCIVELAPTNESYTLSHFNRSYLDDEVSNLHQACLSAFASCIRYPSSDVS 8017  
QY 370 -----DM-----TLK-----I-----WS-MK-Q----- 380  
Db 8018 FIDFDADLVAGLVKPDTKHMQPVEAAKTNRSKSGSDGWSPLQLQIRAAAYSAYSSAP 8077  
QY 381 -----D-----N-----C----- 383  
Db 8078 EDRIIRDTTYIKLGLOISIAIOLANRLKDGLLVQASDVMSPCSELASAVOSRQTPV 8137  
QY 384 -----V-----H----- 385  
Db 8138 LDERGDFEGFKHYRGAALQSHRIATEKVASVRCPTPLQSGMLSEYTHSDGHQYFNHTP 8197  
QY 386 -----D-----LQ-----QH----- 390  
Db 8198 YATEADIDSSKLSAWSKVLECHELLRTGFGVDDHEHPFVMLTYTFDFVIDLEIQASSR 8257  
QY 391 -----N-----K-----E-----I----- 394  
Db 8258 EGSVYSEOKASESVKNLHLPMPWRSLGVBGRCLQPSAHHAIFDAESLRLIMTDLQ 8317  
QY 395 -----Y-----TI-----K-----WS-----P----- 401  
Db 8318 SALSNGVVPRLTIDGALHILNSQADVESQRTFWSQKLSGAPVTFPPNMTVPRISDTE 8377  
QY 402 -----TG----- 403  
Db 8378 AANVELVLYNKRKLEARCOELGVSMQSVQAAWARLLSAYTGESQVTFGVVLGRTSPA 8437  
QY 404 -----P-GTN-----N-----P--NA-N----- 412  
Db 8438 TADAAFPCTITLPSVNTAVDDSQFLKDLMSYNATIQKHQFTPLTIRNVAESTSEALFD 8497  
QY 413 L-----M-----LA-S-----A-S----- 419  
Db 8498 SLFVYQRPMDVVDSSWKIIREKASVELAVSVEMEALSDGLGLRLTIDPAQVPHEQK 8557  
QY 420 -----P-D-----S-----T-----V-----R----- 425  
Db 8558 IMLQMEVMTIAGLKFEDAINTSVMIIIPKPIIATDFPKYLHEMTEASVKSYSRIAME 8617  
QY 426 -----L-----W-----D-----VDRG----- 432  
Db 8618 FVDALDEGGQISSRHWTYRQLDDEANKIAHLIDRGVKPGDIIATISFDKCPASAFYGIL 8677  
QY 433 ----- 433  
Db 8678 KAGCAFCADPTAPAAKAFILEDNARVLLTSDSIRSELRELTOCDIIDLINFNKNEL 8737  
QY 434 -----C-I----- 435  
Db 8738 STSSVPVSLAPSVSIVLYTSQTTGTPKGEITHDNVQVMSFKRLFKGRWTDSESL 8797

QY 436 -----H-----T-----LTK----- 440  
Db 8798 QFASVHFDVSVLEQFQWTWIVGMRLVCAPRDILIEDIAGFLDTMQITHLDTLPSGLRLDLP 8857  
QY 441 -----H-----Q-----E--P--- 444  
Db 8858 ALVPSLHKGVFTTGGESLQDQINTWGDVGCFLNFYGPTECTIGTVTFPCVPKEGKPSNI 8917  
QY 445 -----V--Y-----S--V-----AF----- 450  
Db 8918 GWQFDMVGCYVLAPGTQTPVLGAI GELCISGKLVGKGLNRPDLTADCFPYLDAGFCERV 8977  
QY 451 -----S-----P----- 452  
Db 8978 YRTGDLVRLFHGSDIDFLGRKONQVKLRQORLEIDEBIAVIRKQDIQDTVCIVAKHPKQ 9037  
QY 453 -----D-----G----- 454  
Db 9038 DXDQLIAFIGINESRKQKPELCPAESTRHLIQTABAACEERLPGYMWVTHFLPIQRIPL 9097  
QY 455 -----R-----Y-----LA----- 458  
Db 9098 SVNKKVEEKLRLQYADLPTTVIQTATQADSQQSLDGEQKVAQALAEKLLKIDNDLTP 9157  
QY 459 -----S-G-----S-P-----D 463  
Db 9158 SSNIFSLGSSISAIQFSKKLKASGFTTVQVATVTKNPTISRLTKALATSTGRSGEIA 9217  
QY 464 -----K-----C----- 465  
Db 9218 AKQVISACRQRHMGTVTRVLRCKADDIEAIACTPLQOQIIISRLASESSLYFNSFKFNA 9277  
QY 466 -----W---NT-----V--H-----I-- 468  
Db 9278 QGVDLQKLEGAFNQALERTQILRTFFIEDDGVQAVRKTGHLPMWMTLEVYDLASVDGVF 9337  
QY 469 -----W---NT----- 471  
Db 9338 AKRKQKRSYNTSHLTVPFEIVIVRSGETFVSVDLHLYDGNSTFDILMNVSKLYNSQ 9397  
QY 472 -----Q-----T----- 473  
Db 9398 EADFGKFPVDCLAFGLFRNVQAKQFWDHLDPDKVSASMPPLIDNPASHDVLTCLASLDIL 9457  
QY 474 -----G---ALV-----HSYR---GT-----G-----G-----I 486  
Db 9458 NOADELRSLGVTQALVQATWATVTLRKH-YQGAIGTVVSGRSIDPDGVENVIGPLFNTI 9516  
QY 487 -F--E--V--C-----W----- 491  
Db 9517 PLYLRCEPGDTWQTLVQORCHDFTTALPYQHTPLRDIWKVCKNGHSRPLFDALFVYQGITL 9576  
QY 492 -----N-----AAG-----D----- 496  
Db 9577 DNSDTNHSILKPLEDDSPFADYPLSFEEAAGGNLKIISVAAKASICNETKARELIDEPH 9636  
QY 497 -----K-----VGAS-----A--SD---GS----- 506  
Db 9637 QAFLAMNKSPENEVGASIGHTFPERLARTEDGTYKREIQARDTSDTFHWGSEASVIRSEIATL 9696  
QY 507 --V-----CV-LD-----L--R-K 514  
Db 9697 AGVQERADIDERTSIFEVGLDSDVAVKLSRLK 9728

## RESULT 6

S02392

alpha-2-macroglobulin receptor precursor - human

N:Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor

C:Species: Homo sapiens (man)

C:Date: 14-Aug-1998 #sequence revision 14-Aug-1998 #text\_change 09-Jul-2004

C:Accession: S02392; S30027; I37998; A39210; S12538

R;Herz, J.; Hamann, U.; Rogne, S.; Myklebost, O.; Gausepohl, H.; Stanley, K.K.  
EMBO J. 7, 4119-4127, 1988  
A:Title: Surface location and high affinity for calcium of a 500-kd liver membrane protein  
A:Reference number: S02392; MUID:89210795; PMID:3266596  
A:Accession: S02392  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-4544 <HER>  
A:Cross-references: UNIPROT:Q07954; EMBL:X13916; PIDN:CAA32112.1; PID:G34339  
R;Kristensen, T.  
submitted to the EMBL Data Library, October 1990  
A:Reference number: S30027  
A:Accession: S30027  
A:Molecule type: mRNA  
A:Residues: 3275-3864 <KRI>  
A:Cross-references: EMBL:X55077  
R;Herz, J.; Kowal, R.C.; Goldstein, J.L.; Brown, M.S.  
EMBO J. 9, 1769-1776, 1990  
A:Title: Proteolytic processing of the 600 kd low density lipoprotein receptor-related protein  
A:Reference number: S12538; MUID:90269210; PMID:2112085  
A:Contents: annotation; site of proteolytic cleavage  
R;Kutt, H.; Herz, J.; Stanley, K.K.  
Biochim. Biophys. Acta 1009, 229-236, 1989  
A:Title: Structure of the low-density lipoprotein receptor-related protein (LRP) promotes  
A:Reference number: I37998; MUID:90089395; PMID:2597675  
A:Accession: I37998  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-11 <RES>  
A:Cross-references: EMBL:X15424; NID:G34408; PIDN:CAA33464.1; PID:G34409  
R;Strickland, D.K.; Ashcom, J.D.; Williams, S.; Burgess, W.H.; Migliorini, M.; Argaves,  
J. Biol. Chem. 265, 17401-17404, 1990  
A:Title: Sequence identity between the alpha2-macroglobulin receptor and low density lip  
A:Reference number: A39210; MUID:91009181; PMID:1698775  
A:Accession: A39210  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 150-166;234-238,'X',240-245,'X',247-252,'G',686-695;902-916;1096-1109;'S',17  
C:Genetics:  
A:Gene: GDB:LRP1; APR: LRP; A2MR  
A:Cross-references: GDB:119694; OMIM:107770  
A:Map position: 12q13.1-12q13.3  
C:Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated protein (see PIR:A39875).  
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding  
C:Keywords: beta-hydroxyaspartic acid; beta-hydroxyaspartic acid; calcium binding; glycoprotein  
F:1-19/Domain: signal sequence #status predicted <SIS>  
F:20-3943/Product: alpha-2-macroglobulin receptor 515K chain #status predicted <S15K>  
F:27-64/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F:72-108/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F:115-148/Domain: LDL receptor ligand-binding repeat homology <EG1>  
F:154-188/Domain: EGF homology <EG2>  
F:198-239/Domain: LDL receptor YWTD-containing repeat homology <YW01>  
F:240-281/Domain: LDL receptor YWTD-containing repeat homology <YW02>  
F:292-334/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
F:335-378/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
F:379-420/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
F:421-468/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
F:478-519/Domain: EGF homology <EG3>  
F:571-613/Domain: LDL receptor YWTD-containing repeat homology <YW07>  
F:614-659/Domain: LDL receptor YWTD-containing repeat homology <YW08>  
F:660-710/Domain: LDL receptor YWTD-containing repeat homology <YW09>  
F:711-752/Domain: LDL receptor YWTD-containing repeat homology <YW10>  
F:753-799/Domain: LDL receptor YWTD-containing repeat homology <YW11>  
F:807-842/Domain: EGF homology <EG4>  
F:854-890/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
F:895-931/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
F:936-971/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
F:976-1011/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
F:1015-1051/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
F:1062-1097/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F:1104-1140/Domain: LDL receptor ligand-binding repeat homology <LDL9>  
F:1145-1182/Domain: LDL receptor ligand-binding repeat homology <LDLA>

F:1185-1221/Domain: EGF homology <EG5>  
F:1227-1261/Domain: EGF homology <EG6>  
F:1269-1308/Domain: LDL receptor YWTD-containing repeat homology <YW12>  
F:1309-1355/Domain: LDL receptor YWTD-containing repeat homology <YW13>  
F:1356-1398/Domain: LDL receptor YWTD-containing repeat homology <YW14>  
F:1399-1445/Domain: LDL receptor YWTD-containing repeat homology <YW15>  
F:1446-1488/Domain: LDL receptor YWTD-containing repeat homology <YW16>  
F:1489-1531/Domain: LDL receptor YWTD-containing repeat homology <YW17>  
F:1540-1578/Domain: EGF homology <EG7>  
F:1583-1626/Domain: LDL receptor YWTD-containing repeat homology <YW18>  
F:1627-1669/Domain: LDL receptor YWTD-containing repeat homology <YW19>  
F:1670-1713/Domain: LDL receptor YWTD-containing repeat homology <YW20>  
F:1714-1753/Domain: LDL receptor YWTD-containing repeat homology <YW21>  
F:1754-1796/Domain: LDL receptor YWTD-containing repeat homology <YW22>  
F:1797-1846/Domain: LDL receptor YWTD-containing repeat homology <YW23>  
F:1850-1886/Domain: EGF homology <EG8>  
F:1934-1976/Domain: LDL receptor YWTD-containing repeat homology <YW24>  
F:1977-2019/Domain: LDL receptor YWTD-containing repeat homology <YW25>  
F:2020-2063/Domain: LDL receptor YWTD-containing repeat homology <YW26>  
F:2064-2105/Domain: LDL receptor YWTD-containing repeat homology <YW27>  
F:2106-2151/Domain: LDL receptor YWTD-containing repeat homology <YW28>  
F:2159-2194/Domain: EGF homology <EG9>  
F:2193-2241/Domain: LDL receptor YWTD-containing repeat homology <YW29>  
F:2253-2294/Domain: LDL receptor YWTD-containing repeat homology <YW30>  
F:2344-2388/Domain: LDL receptor YWTD-containing repeat homology <YW31>  
F:2389-2429/Domain: LDL receptor YWTD-containing repeat homology <YW32>  
F:2430-2473/Domain: LDL receptor YWTD-containing repeat homology <YW33>  
F:2482-2517/Domain: EGF homology <EG10>  
F:2524-2561/Domain: LDL receptor ligand-binding repeat homology <LDLB>  
F:2566-2600/Domain: LDL receptor ligand-binding repeat homology <LDLC>  
F:2605-2639/Domain: LDL receptor ligand-binding repeat homology <LDLD>  
F:2652-2688/Domain: LDL receptor ligand-binding repeat homology <LDLE>  
F:2695-2730/Domain: LDL receptor ligand-binding repeat homology <LDLF>  
F:2734-2769/Domain: LDL receptor ligand-binding repeat homology <LDLG>  
F:2774-2812/Domain: LDL receptor ligand-binding repeat homology <LDLH>  
F:2818-2853/Domain: LDL receptor ligand-binding repeat homology <LDLI>  
F:2858-2897/Domain: LDL receptor ligand-binding repeat homology <LDLJ>  
F:2904-2939/Domain: LDL receptor ligand-binding repeat homology <LDLK>  
F:2944-2980/Domain: EGF homology <EG11>  
F:2986-3021/Domain: EGF homology <EG12>  
F:3029-3068/Domain: LDL receptor YWTD-containing repeat homology <YW34>  
F:3069-3113/Domain: LDL receptor YWTD-containing repeat homology <YW35>  
F:3114-3156/Domain: LDL receptor YWTD-containing repeat homology <YW36>  
F:3157-3200/Domain: LDL receptor YWTD-containing repeat homology <YW37>  
F:3201-3241/Domain: LDL receptor YWTD-containing repeat homology <YW38>  
F:3242-3284/Domain: LDL receptor YWTD-containing repeat homology <YW39>  
F:3294-3330/Domain: EGF homology <EG13>  
F:3334-3369/Domain: LDL receptor ligand-binding repeat homology <LDLL>  
F:3374-3408/Domain: LDL receptor ligand-binding repeat homology <LDLM>  
F:3413-3448/Domain: LDL receptor ligand-binding repeat homology <LDLN>  
F:3453-3489/Domain: LDL receptor ligand-binding repeat homology <LDLO>  
F:3494-3531/Domain: LDL receptor ligand-binding repeat homology <LDLP>  
F:3536-3570/Domain: LDL receptor ligand-binding repeat homology <LDLQ>  
F:3575-3609/Domain: LDL receptor ligand-binding repeat homology <LDLR>  
F:3613-3647/Domain: LDL receptor ligand-binding repeat homology <LDLS>  
F:3654-3690/Domain: LDL receptor ligand-binding repeat homology <LDLT>  
F:3695-3731/Domain: LDL receptor ligand-binding repeat homology <LDLU>  
F:3741-3776/Domain: LDL receptor ligand-binding repeat homology <LDLV>  
F:3785-3822/Domain: EGF homology <EG14>  
F:3828-3860/Domain: EGF homology <EG15>  
F:3868-3911/Domain: LDL receptor YWTD-containing repeat homology <YW40>  
F:3913-3969/Domain: LDL receptor YWTD-containing repeat homology <YW41>  
F:3944-4544/Product: alpha-2-macroglobulin receptor 85K chain #status predicted <EX>  
F:3944-4544/Domain: 85K chain extracellular #status predicted <EX>  
F:4013-4056/Domain: LDL receptor YWTD-containing repeat homology <YW42>  
F:4057-4099/Domain: LDL receptor YWTD-containing repeat homology <YW43>  
F:4100-4142/Domain: LDL receptor YWTD-containing repeat homology <YW44>  
F:4151-4182/Domain: EGF homology <EG16>  
F:4200-4231/Domain: EGF homology <EG17>  
F:4236-4267/Domain: EGF homology <EG18>  
F:4272-4303/Domain: EGF homology <EG19>  
F:4308-4339/Domain: EGF homology <EG20>

F;4344-4374/Domain: EGF homology <EG21> F;4377-4408/Domain: EGF homology <EG22> F;4421-4444/Domain: transmembrane #status predicted <TM> F;4445-4544/Domain: intracellular #status predicted <INT> F;166,2998/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted F;2958/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted F;4075,4125,4278/Binding site: carbohydrate (Asn) (covalent) #status predicted	Query Match Best Local Similarity 70.5%; Score 2580.5; DB 1; Length 4544; Matches 434; Conservative 61; Mismatches 12; Indels 3805; Gaps 385;
QY 1 MS-----I-----SSDE-----V-----N-F-L-----11     	
DB 88 MSRLCNGVQCMDSDEGPHRCRLQGNCSRLGQHCHVPTLDGFTCYCNSSFOLOADGKT 147	
QY 12-----VY-----R-----Y-LQ-----E-----18     	
DB 148 CKDFDECSVYTCSQLCTNTDGSFICGVBEGLLQPDNRSCAKNEPDRPPVILLIANSQ 207	
QY 19-----SG-----FS-----H-SA-----F 26     	
DB 208 NILATYLSGAQVSTIPTSTROTAMDFSYANETVCWVHVGDSAAQTOLKCARMPGLKGF 267	
QY 27-----TFGIK-S-HISQ-----S-N-----I-NG-----41     	
DB 268 VDEHTINISLHHVEQMAIDMTGTFYFVDDIDRIFVCNRNGDTCVTLLDLELYNPKG 327	
QY 42 -AL-----VP-----P-----A-A-48     	
DB 328 IALDPAMGVFFTDYQIPKVERCDMDGQNRKTLVDSKIVFPHGITLVLRSRLVYWADAY 387	
QY 49 -LI-----SLIQK-L-Q-Y--V-E-----A-EVS-I-----N 67     	
DB 388 LDVIEVVDGKGRQTHIQ-GILIEHLYGLTVFENLYATNSDNAQAQKTSVIRVNRFN 446	
QY 68 -E-----D-GTL-----P-----72     	
DB 447 STEYQVTVRVDKGGALHIYHQRRQPRVSRSHACENDQYKPGGSDICLLANSHKARTCR 506	
QY 73 -F-----DG-----R-----PIE-----SL--81     	
DB 507 RSGFSLGSDGSKCKPEHELFLVYKGRPGIIRGMDMGAKVDPDEHMIPIENLMNPRALDF 566	
QY 82-----S-LI-----D-----AV-M-----PD--90     	
DB 567 HAETGFIYFADTTSYLIGROKIDGTERETILKDGHNVEGVAVDWMDNLYWTDGPKKT 626	
QY 91 V-Y-----QTR-----QO-----97     	
DB 627 ISVARLEKAAQTRKTLIEGKQWTHPRALVVDPLNGWYWTWDEEDPKDSRRGRLEAAMWG 686	
QY 98-----ISVARLEKAAQTRKTLIEGKQWTHPRALVVDPLNGWYWTWDEEDPKDSRRGRLEAAMWG 102     	
DB 687 SHRDI FVTSKTLWPNGLSLDIPAGRLYWDYAFYDRIETILLNGTDKIVYEGPELNHAF 746	
QY 103-----L-----L-----AQQQA-----109     	
DB 747 GLCHGNLYFWTYRSGSVYRLRGVGGAPPTVTLRSRPPPIFEIRMYDAQQQVGTNK 806	
QY 110-----AA-----A-A-----A-----A-AASQ--120     	
DB 807 CRVNGGCSLCLATFGSRQCAEAOVDLDAGVTCLANPSYVPPQCPQGEFACANSRC 866	
QY 121 Q-G-----S-----A-----K-----N-G-----E 128     	
DB 867 IQERWKCDGNDCLDNDSEAPALCHOHTCPSDRFKCENNRNRCIPNRLWCDGNDGNSDE 926	
QY 129 -N-T-A-----N-----G-----E-----E-----N 136     	
DB 927 SNATCSARTCPNPQFSCASGRCIPISWTCDLDDCGDRSDSASCAYPTCFPLTQTCNN 986	
QY 137 G-----A-----H-----T-----I-----141     	



Db 2065 KLYWCDARTDKIERIBLETGENREVLSSNNMDMFSVSVFEDFIYWSDRTHANGSIKRG 2124  
QY 297 -----TG----- 298  
Db 2125 KONATDSVPLRTIGVQLDKIVFNDRDRQKGTWCAVANGCCOQLCYRGRGORACAH 2184  
QY 299 -----E-A-----K-----Q----- 304  
Db 2185 GMLAEDGASCREVAGYLLYSERTILKSIHLSDBERNLNAPOPPEDPEHMKVIALAFDYR 2244  
QY 305 -----P-----F-----H-----S-----A----- 309  
Db 2245 AGTSPGTNRIFFSDIHFGNIOQINDGSRRIITIVENVGSVEGLAYHRGMDTLTYSTT 2304  
QY 310 -----P-AL-----D-----V-D-----W-Q-----S 318  
Db 2305 SIITRHTVDTQTRCAFERETIVTMSGDDHPRAVFLDECONLMFTWNEOHPSIMRAALS 2364  
QY 319 -N-----N-----T----- 321  
Db 2365 GANVLTLIEKDITPTNGLAIDHRAEKLYPSDATLTKIERCEYDGSYHYILKSEPVHFG 2424  
QY 322 -----F-----A-----SC-S----- 326  
Db 2425 LAVYGEHIFWTDWVRRAVORANKHVGSNMKLRLVDIPOOPMGIIAIVANDTNSCELSPCRI 2484  
QY 327 -----T-----DM-----CI----- 331  
Db 2485 NNGGCDLCLLTHQGHVNCGRGRILQDDLTCTRAVNSSCRAQDEPECANGECINFSLTC 2544  
QY 332 -H-----H-----V-----C-K----- 335  
Db 2545 DGVPHCKDKSDEKPSYCNRRCKKTRQCSNGRCVSNMLWCNCGADCGDGSDEIPCNKTA 2604  
QY 336 -----Q-----IG----- 338  
Db 2605 CGVGEPRCRDGTICGNSSRCNQFVDCEDASDEMNCSATDCSSYFRGLGVKGLVFPQCERTS 2664  
QY 339 -----D-----RP-----I-----K----- 343  
Db 2665 LCVPASWCDGANDCGDYSDERDCPVKRPCLPNYFACPSGRCI PMSWTCKDEDDCEHG 2724  
QY 344 -T-----F-----Q-----G-----H----- 349  
Db 2725 EDETHCNKFCSEAFQCNHRCISKQWLCGSDGDDCGSDEAAHCRGKTCGPFSSFCPT 2784  
QY 350 -----N-----EV-----N-A-I-K----- 356  
Db 2785 HVCVPERWLCDDGKOCADGADESIAAGCLYNSTCDDREFMCONRQCIPKHFVCDHDDCA 2844  
QY 357 -----W-----D-----P-T----- 360  
Db 2845 DGSDESPECEYPTCGSPBFCANGRCCLSRQWCEGENDCHDQSDAPKNPHCTSPHCK 2904  
QY 361 -----G-----N-----L----- 363  
Db 2905 NASSQFLCSSGRVABALLCNGQDDCGSDSDRGCHINECLSKLSCGQDCEDLKIGFK 2964  
QY 364 -----L-----A-----S-----CS-----D----- 369  
Db 2965 CRCRPGFLKDDGRTCADVDECSSTTPPCSQRINCINTHSYKCLCEVAGPRGGPHSKAV 3024  
QY 370 -D-----M-----T-I-K-----I-W-----SM-- 378  
Db 3025 TDEEPFLIFANRYLRKLNLDGNSYTLKQGLNNAVALDFDYREOMIYWTDTTQGSMTIR 3084  
QY 379 -K-----Q-----D-----N-----C----- 383  
Db 3085 RMHLNGSNVQVLRHTGLSPDGLAVDWVGNLYWCDKGRDTIEVSKLNGAYRTLVSSGL 3144  
QY 384 -----V-----H----- 385  
Db 3145 REPRALVVDQNGLYWTDWGDHSLIGRIGMDSSRSRVIVDTKI TWPNGLTLDYVTERIY 3204

QY 386 -----D-----L-----Q-----H-----N-----K----- 392  
Db 3205 WADAREDIYEFASLDGSRHVVLSQDI PHIFALTLPEDVYVWTDWETKSNRAHKTGTN 3264  
QY 393 E--I----- 395  
Db 3265 KTLILSTLHRPMDLHVFPALRQPDVFNHPCKVNVNGGCSNLCLLSPGGGHKACACTPNFYLG 3324  
QY 396 -----T-----K-W-----S-----P-----TG 403  
Db 3325 SDGRTCVSNCTASQFVCKNDKCI PFWMKCDTDDCGDHSDEPPDCPEFKRPGQFCSTG 3384  
QY 404 -----P-----G-----TN-----N----- 408  
Db 3385 ICTNPAFICDGDNDQDNDSEANCDIHVCLPSQFKCTNTNRCIPGIFRCNGQDNCGDGED 3444  
QY 409 -----PN-----ANL-M----- 414  
Db 3445 ERDCPEVTCAPNQFQCSITKRCIPRVWVCDRNDVCGSDPEANCTQMTGCVDFRCKDS 3504  
QY 415 ---L-A----- 416  
Db 3505 GRCIFARMKCDGEDDCGSDGSDPEKBECDERTCEPYQFRCKNNRCVPRWQCYDNDCGDN 3564  
QY 417 -----S-----A-S-----F----- 420  
Db 3565 SDEBSCTPRPCESEBFCANGRCIAGRWKCDGDHDCADGSDKDCPTPRCDMDQFQCKSGH 3624  
QY 421 -----D-S-----T-VR-----L-W-----D 428  
Db 3625 CIPLRWCDADADCMGSDDEBACGTGVRTCLPDEFCNNNTLCKPLAWKCDGEDDCGSDNSD 3684  
QY 429 -----V-----DRGIC--I-----HTLTKH- 441  
Db 3685 ENPECARFVCPNPPFRCKNDR-VCLMTGROCDGTNCGDGTDEBDCBPPTAHT-T-HC 3741  
QY 442 -----Q-----E-P-----V-- 445  
Db 3742 KDKKFLCERNQRLSSLRCNMFDDCGSDGSDDEDCSIDPKLTSCATNASICGEARCVRT 3801  
QY 446 -Y--S--V-----A--F----- 450  
Db 3802 EXAAVCACRSQGHPTVPGQPCQDINECLRFGTCSQLCNNTKGHLCSGARNFMKHTNCK 3861  
QY 451 -----S-----P-----D-----GR----- 455  
Db 3862 AEGSEYQVLYIADDDNEIRSLFPFGHPHSAYEQAFQGDSEVRI DAMDVHVKAGRYVTNWHT 3921  
QY 456 ---Y--L--A--S-----G-----S-----F-----D- 463  
Db 3922 GTISTRSLPPAAPPTTSNRRHQIDRGVTHNLISGLKMPRGIAIDWVAGNVYVWTDSDGRV 3981  
QY 464 -----K-C-----V--H-I-----W-N-----T-- 471  
Db 3982 IEVAQMKGNRKTLSGMIDEPHAIWVDPLRGTYWSDMGNHPKIEATAAMDGLRLETQV 4041  
QY 472 ---Q--TGAL-V--HS--Y-----R--GT-----G----- 484  
Db 4042 DNIQWPTG-LAVDYHNERLYWADAKLSVITGSRNLNGTDP IVAADSKRGLSHPSIDVFED 4100  
QY 485 ---G-----I--F-----BV-----C-W-- 491  
Db 4101 YIYGVTYINNRFKTHKHGHSPLVNL TGGLSHASDVVLVHQHKQEPVTNPPCRKKEWLC 4160  
QY 492 -----N-----A----- 494  
Db 4161 LLSPSGPVCTCPNGKRLDNGTCVPVSPPTPPDA PRPGTCNLQCFNGGSCFLNARQPKC 4220  
QY 495 -----GDK-----V--G--A-- 500  
Db 4221 RQCPRYTGDKCBELDQCBWHCRNGGTCAASPSGMPTCRPTGTGPKCTQQVCAGYCANNS 4280

QY 501 -----S-----ASDGS----- 506  
Db 4281 TCTVQGNQPCRCPLPGFLGDRCOYRQSGCYCENFGTCQMAADSGRCRCCTAYPEGSRCE 4340  
QY 507 V-----CVL-----DL-----R-----K 514  
Db 4341 VNKSRCLGACVVKQSGDVTCTDGRVAPSLCTCVGHCSNGSGTMSK 4392

RESULT 7  
T13931  
projectin - fruit fly (Drosophila melanogaster) (fragment)  
C:Species: Drosophila melanogaster  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T13931  
R:Daley, J.; Southgate, R.; Ayme-Southgate, A.  
J. Mol. Biol. 279, 201-210, 1998  
A:Title: Structure of the Drosophila projectin protein: isoforms and implication for pro  
A:Reference number: 217815; MUID:98300339; PMID:9636710  
A:Accession: T13931  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-6658 <DAL>  
A:Cross-references: UNIPROT:O76281; EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AA02  
C:Genetics:  
A:Gene: projectin  
A:Cross-references: FlyBase:FBgn0005666  
A:Map position: 4  
A:Note: intron positions not resolved (incomplete sequence)  
C:Keywords: muscle

Query Match 70.5%; Score 2580.2; DB 2; Length 6658;  
Best Local Similarity 8.8%; Pred. No. 3e-52; Indels 4578; Gaps 400;  
Matches 446; Conservative 55; Mismatches 10;

QY 1 M-S-----IS-----S-----D-----EVN----- 9  
Db 674 LPSSEIKNIPYNTKISIIETVRKHTGIYKIIAVNEHGQDEATVEVNILAPPSKPRGPLD 733  
QY 10 -----F-----L-VY-----RY----- 19  
Db 734 VKDVTDSCKFKWKPKPEDDGGKPIISAYQVEKFKDKQGRWVPLGRTSANDTEFDVKGLQE- 792  
QY 20 G-----F-----S-----H-----S 24  
Db 793 GHEVQFRVKAINEEGESDPLSDSDSIIAKNPYDAASKPGTPIVNDVNEHVMVKLWEAPRS 852  
QY 25 -----A-----F-----T-----P-----GI-K 31  
Db 853 DGGAPISGIYIIKKDKFSPIDWDEILSTNTSVPATVEGLVEGNIYQFRVAVNKAQFSDP 912  
QY 32 S-----H-----I-----I-----S 35  
Db 913 SDATEPLAKPRNLKPYINRDKMKPIKVRAGQPVKFDVDVKGEPAPSLTWFLKETELTST 972  
QY 36 -----OS-----NING-----A-----L-V- 44  
Db 973 GQVRLENIDYNTKLTLLDTRKQSGQYKLAENINGVDEAVVEVILDKPSKEGPLEVS 1032  
QY 45 -----P-----P-AAL-I-----SI-----IQ-KGL-----Q 58  
Db 1033 DIHKEGCKLWRKPKODGGIPITGYVIEKMDTATGKWPAGSVDPKDYDIEIKGLDPNHR 1092  
QY 59 Y-----V-----E-----A-----E-V----- 64  
Db 1093 YQFRVAVNEEGESEPLETESAITAKNPFVDSAPPGLPELEDWDEHHVKLWKEPPTRDGG 1152  
QY 65 S-I-N-----E-----D-G-----T-----L----- 72  
Db 1153 SPITNVIIEVMDKDPGEFVKAVETDSPVCKGVVKLEEGQYKFRVAVNKAQSPDSEQ 1212  
QY 73 -----F-----D-----G-----R-----P-IE-----S 80

Db 1213 TNWHVAKPRFLKPHIDRVNLKPVIVKTLGLSILDIRGEPAPKVEFFNNSSVTSDEHS 1272  
QY 81 -----L-----S-LI-----D-A-----V----- 87  
Db 1273 VKIDNDVYNTKPFVWRAORSQSGYIIKATNEVGEDEALEVTLVGLGPKGPKGLQVNDI 1332  
QY 88 -----MPD-----V-----V-----Q----- 93  
Db 1333 TKHSCCKLWKEKPDGSGPIDYIEIEKLDPHTCQWLPCGSKSTPEAKVIGLHKGKAYKFR 1392  
QY 94 -----T-----R-----Q 96  
Db 1393 VRVAVKEGESEDLTEKPIIAKNPYDEPDRPGKPEPTNWDKOFVDLAWDPKNDGGAPIQ 1452  
QY 97 -----Q-----A-----Y-R-----D-K- 102  
Db 1453 KYVIQMRDKSGRAWDSATVPDGKNGTGTGVBEGHEYEFRIIVAVNKAQSPDSDVSKSV 1512  
QY 103 LA-----Q-----Q----- 106  
Db 1513 IAKPRFLKPHIDRKNLQKKIMRSGQMLHIDALIKAPPAKVTVTYNKTETKSDHIKIE 1572  
QY 107 -----QA-----A-----A-----A----- 111  
Db 1573 EDYKTTFIMPVKRADRGYIIVTAKNDSGSDTVEVEVLCKPSKPKGLAVSNVTAETL 1632  
QY 112 -----A-----A-----A-----A----- 113  
Db 1633 HLKWEKPEDDGGDPIBOYLVERMDTETGRWVPLTKTTEADVTGLTEGKEYLFRVKA 1692  
QY 114 -----A-----AA-----A-----A----- 119  
Db 1693 SEGESEPLVTDIPTKAKNPFDAADTPGKQIVDWSGNHCDLKWRAPEDDGGASITGYVE 1752  
QY 120 -----Q-----Q-----Q-----G-S-----AK- 125  
Db 1753 RKDPNTGKOKALETSTPDKARVNDLIAGNKYQFRIMAVNKAQSKPSPSQMTAKDR 1812  
QY 126 -----N-----GE-----N----- 129  
Db 1813 FAPPKIDRTNIKDITSKAGQIRFDIKVSGEPPATKVLHKNKARLENDNSNYNIDMESYR 1872  
QY 130 -----T-A-N-----G-E-----E----- 135  
Db 1873 TKLTVPISKRFHSGKYTLKAENESGRDEASFEVIVLDKPGPEGLRVTDVHKEGCKLKW 1932  
QY 136 N-----G-AH-TI-----A-NN-----H-----T- 146  
Db 1933 NAPLDDGGLPIDHYIIEKMDVESGRWLPSPGRFKESFAELNNLEPSHEYKFRVLA 1992  
QY 147 -----D-M-M-E-----VD----- 152  
Db 1993 SBPLTGEQSVIAKNPDEPGKPGTPEAVDWDKHDVLDVWRPPINDGSGSPITGYVVEKREK 2052  
QY 153 G-D-V-----E-IP-----P-N-----KAV----- 163  
Db 2053 GTDKMKGTEITIPCLGEBCKATVPTLNCEVEFRVKAINAAGPGEPSDASKPIITKPR 2112  
QY 164 -----VL-----R-----G-----H-E----- 169  
Db 2113 KLAPTILDPKIRTYNFKSGSPIFLDINISGEPADVTNQNKNKSVQTSFISHIENLPY 2172  
QY 170 -S-----E-----V-F-----I-----C----- 175  
Db 2173 NTKYINNPERKDTGLYKISAHNFYQDQVFEQIINIITKPGKPGPLEVSEVHKCGCKUK 2232  
QY 176 -----AW-----N-----P-----V----- 180  
Db 2233 WKKPKDDGGEVSVYLVKFDPDGTGLVLPVGRSDGPEYNVDGLVPGHDYKFRVKA 2292  
QY 181 -----S-----D-L-----L-----AS-----GS----- 188  
Db 2293 ESEPLETLGSIIAKOPFSVPTKPGVPEPTDWTANKVELAWPEPASDGGSPIQGYIVEVKD 2352

QY 189 -----G-----D-S-TA-----193  
Db 2353 KYSPLWEKALETNSPTTATVQGLIEGNEYQFRVALNKNLSESPDKIFTAKRYIA 2412  
QY 194 -----R-----I-----W-----N-----L-198  
Db 2413 PKIDRNLRLNITLSSGALKLDANITGQAPKPVEMKLSNYHLQSGKNVTIETPDYTKLV 2472  
QY 199 -----S-E-----N-----S-----T-----S 204  
Db 2473 IRPQRTDSGEYLVTANTSGKSVLNVVVITDKPSPNGPLOISDVHKEGCHLWKRPS 2532  
QY 205 -----G-----ST-Q-----L-----V-----L 211  
Db 2533 HDGGTPIEYQIDKLEPETGCMIPSCRESTAPOVDVTGLSPGNEYKFRVSANAGESQPL 2592  
QY 212 -----R-----H-----C-I-----215  
Db 2593 VGDESIVARNPFDEPGKPNLKATDWDKHVDLAWTPPVIDGGSPISCVIIIEKQDKYKGM 2652  
QY 216 -R-----EG-----218  
Db 2653 ERALDVPADOCKATIPDLVEGQYKFRVSANAGTGEPSDSTPPIIAKARNKPIIDRS 2712  
QY 219 -----GQ-----D-V-----P-----S-----N-K-----226  
Db 2713 SLVEVRIKAGQSFTFDCKVSGEPAPQTKMLLKKEVYKDNVNTVNTKLKVSATR 2772  
QY 227 -----D-VT-----S-----230  
Db 2773 SDSGIVTVAENANGESADVKVTVIDKPAPPNGPLKVDEINSECTLHWNPPDDGGOP 2832  
QY 231 -----LD-----W-----N-----SE-----236  
Db 2833 IDNVVVKLDETTGRWMTAGETDGPVTKLVKGLTFCHKYKFRVRAKNRQGTSEPLTTAQ 2892  
QY 237 -----G-TL-----L-----A-----TG-----243  
Db 2893 AIIAKNPFDPVTPKPTTKDFKDFEVDLEWTPREADGGSPITGYVVEKDKFSPWEKC 2952  
QY 244 -----S-----Y-----245  
Db 2953 ABISDDITIAHVLDLIEGLKYBFRVAVNKAAGPGSPDATETHVARPKNTPPKIDRNEMS 3012  
QY 246 -----D-----G-----F-----248  
Db 3013 DIKIKAGNVFEDVPVTGPEPLPSKDWTHEGNMIINTDRVKISNFDRTKIRILSATSDTG 3072  
QY 249 -----AR-----I-----W-----T-----253  
Db 3073 VYTLTARNINGTRHNKVTILDAPSPPEPALRNGDVKNKSVLVRWPPKDDGGSITHY 3132  
QY 254 -----K-D-----G-----NL-----A-----S-T-----261  
Db 3133 VVEKMDNEAMRWVPVGDCTDEIRADNLNENHDYSFRVAVNKGQSQPLTTSQPIATKD 3192  
QY 262 -L-----GQ-----H-----K-G-PI-----F-----A-----271  
Db 3193 PVSHDPKQGPQATDNGKHFDVLEWSTPKRDGGAPISVYIIIEKRPFGQWERAUVGLDN 3252  
QY 272 -----L-----K-----273  
Db 3253 CKAHVPELTNGGEYEFVAVNRGGSPDSPSTIICKPRFLAPFEDKSLNDITVHAG 3312  
QY 274 -----W-N-KK-----G-N-F-I-L-SA-----284  
Db 3313 NGLGWTLPTEASPRPLITWLYNGKEISNRSGSLGFQNELTEIYVSSLSRSABGRVTLIL 3372  
QY 285 -----G-----V-----D-K-T-----T-----II-----292  
Db 3373 KNEHGSFASAHATVLDPRSPKGPLDITKITRDGCHLTWNVPDDGGSPILHYIIIEKMD 3432

QY 293 -----W-DA-----H-T-----GE-----AK-----301  
Db 3433 LSRSTWSDAGMSTHIVHDVTRLVHRKEYLFRVKAVNAIGESDPLEAVNTIIAKNEFDEPD 3492  
QY 302 -----Q-----302  
Db 3493 APGKLIITDWRDHDIDLOWAVPKSGGAPISEYIIQKKEKSPYTWNRVHPSNKNNTTI 3552  
QY 303 -----Q-F-----P-----PH-----307  
Db 3553 PELTEQGEYEFVAVNQAGQSEPSDMMIRKPRYLPKIIITPLNEVRKICGLIFHTD 3612  
QY 308 -----308  
Db 3613 IHFIGEPAPEATWTLNSNPLLSNDRSTITSIGHHSVVHTVNCQSDSGIVHLLLRNSSI 3672  
QY 308 -----308  
Db 3673 DEGSFELVLDPRGPPGPMYEEITANSVTISWKPPKONGGSEISSYVIEKRDLTHTGG 3732  
QY 308 -----SA-----P-----A-----L-D-----V-----314  
Db 3733 WVPANVYSKYNHAVVPRLLEGTMYELRVMAENLQGRDPLTSDQPVVAKSQYTVPGAP 3792  
QY 315 -----D-----W-Q-SN-----NT-----321  
Db 3793 GKPELTDSDKNHITIKWKQPIISNGGSPIIIGYDIERRDVNTGRWIKINGQVPVTAEQDDR 3852  
QY 322 -----F-----322  
Db 3853 VTSNHQYQYRISAVNAAGNKTSEPSAIFNARPLREKPRFYFDGLICKRIKVRAGEPVNL 3912  
QY 323 -----A-----SC-----325  
Db 3913 NIPISGAPTIEWKKGDKLBEGKRISYETNSERTLFRIDDSNRDRSGKYTVTAANEFG 3972  
QY 326 -----S-T-DW-----329  
Db 3973 KDTADIEVIVDKSPPEGFLSYTETAPDHSILHWYSPKDDGSDITGYIIIEFTEFGVDD 4032  
QY 330 -----C-----330  
Db 4033 WKPVPCTCNTFTVKNLVEGKKYFPRIRAEINYGASEALEGKPLVKSPDPFGAPSQP 4092  
QY 331 -I-----H-----V-----C-----334  
Db 4093 TISAYTPNSANLEWHPPDDCGGKPIGTGYIVERRERGGEWIKCNYPTPTNTSYVSNLRDG 4152  
QY 335 -----K-L-----G-----Q-----DR-----P-----341  
Db 4153 ARYEFRLAVNEAGPGHPKSPDPMPTAEHQRYRPPDPPEPKPDRIITNGVTLWSRPPRTD 4212  
QY 342 -----IK-----T-----F-----345  
Db 4213 GKSRIKGYVEMRPNKNGDKWKTVDIPINSTVYTVPSLKEGEYSFRVVAENVEGRSDPS 4272  
QY 346 -----O-----G-----347  
Db 4273 KPSQPTIBEQNKPCMELGKVRDIVCRAGDDFHSIHVPYLAAPPKNAFWYSNDNMLDDNN 4332  
QY 348 -----H-----348  
Db 4333 RVHKHLTDDAASVVVKNKSRDSSGQYRLOFKDTSFGDTATINVRVLDSPPTLRADVF 4392  
QY 349 -----T-----N-E-----351  
Db 4393 SGDSLTYWNPNDGGSIAQNYIIIEKKEARSSTWSKVSFCTVLFVRIRNLVLNKEYDF 4452  
QY 352 -V-----N-----A-----I-K-----W-----D-P-----359  
Db 4453 RVIAENKQSGDPANTSEPIARHPDIPNTPGIPIHGIDSTEDSTIATMTKPKHGGSP 4512  
QY 360 TG-----NL-----L-----A-----S-----366

Db 4513 TGYIIKRLSDDKWKAVHALCPDLCKIFNLNLIENAEYERVAAVNAQSGSSDL 4572  
QY 367 -C-----SD-----DMT-----LKI-----W----- 376  
Db 4573 IFCRRPHAPKITSDLSIPDMTIVAGDEFRITVPHASPRPTASNGLEVIPGERIKF 4632  
QY 377 -SM-----KOD-----N-----CVH-----D-L-Q----- 388  
Db 4633 DSNDAWYKSAKRDGTSYITLTNNKSGDASC-HVTVDRLPPQGLNAYDITP 4691  
QY 389 -----W-----Q-----H----- 390  
Db 4692 DTCTLAWKTPLDDGSPITNYVVEKLDNCGSWKISSFVNTHYDVMGLEPHYKYNFRVR 4751  
QY 391 -N-----K-----E-----IYT----- 396  
Db 4752 AENQYGLSDPLDIIEPMVAKHQFTVPDEPGQKVIDWDSGNVTLIWTRPLSDGSGRIQGY 4811  
QY 397 -IK-----W-----NP-----SPT----- 402  
Db 4812 QIEYRDILNDSWWAYDYIIKTKYQLYNLNGSEYEFRIKAKNAAGLSKPSPSLRFLK 4871  
QY 403 -G-----PG-----T-----N-----NP-----N 410  
Db 4872 KGKFTVPSPPGAPQVTRGVKNYVDLKWKPLRDGGSRIITGVIIERRDIGGAVVWKNDYN 4931  
QY 411 -----ANL-M-----LA-SA-----S----- 419  
Db 4932 VLDTEYTMNLIEMGDYEFVRFAVNASGRSEPSLCTMPIKVCYGLGKPKDWITRLQDKV 4991  
QY 420 -F-D-----S-TVRLM-----DV-D-R----- 431  
Db 4992 APFGKDYTLQCAASGKSPSTAR-WLRNGKSIQMGEMTCDKDKGVFLHISNVQTDG 5050  
QY 432 -----G-I-C----- 434  
Db 5051 DYTCEAMNSLGFVNTSGYLKIGSPIINRCPSELKPEGDNSKIKIFYSGDQPLTVILKKN 5110  
QY 435 -I-----H-----T-----LT-K-H----- 441  
Db 5111 NEVISDNDTHVKVIFDDYVAIYIANIVKSGDGPYQIEFTNESGSGATGEFYVHTGMP 5170  
QY 442 -----QEP-----V-Y-----S-----V-----AF- 450  
Db 5171 SAPTGMGISYINKNSCMLNWRPPYDGLKVSHYVTERKDVSSPHWITVSSCKOTAFN 5230  
QY 451 -----SP----- 452  
Db 5231 VOGLIENQEIYFRMAVNMENGMGPPLGELNPIRAKDPIDPPSPPAVLKSLRSBETLCNPE 5290  
QY 453 -----D-----GR-Y----- 456  
Db 5291 WKPESDGGAHIQGYWIDKREVSCKMAGVNATICAANQINCINLIEGROYEPRIPAQNV 5350  
QY 457 -L-A-----S-GSF-----D-K-----CV-H-----I-W----- 469  
Db 5351 AGLSTAIVRTSSQDNRSCTGGFASIDCETATHANCINQHNAAFTCTIPGVPKPTISWYKG 5410  
QY 470 -----N-----TQ-T-----GA----- 475  
Db 5411 ARBISNGAGITCTQKVTSLPKYYDVFGEDEAYVCRAVNAKAGAKSTRATLAIMTAPKLV 5470  
QY 476 -----L-----VH-----S-Y-----R- 481  
Db 5471 PPRFRDTAYDKGENVVIKPIFTGLPKPRIHWKGLNIESGGHYTVYVKEKHAVALIIRD 5530  
QY 482 G-----TG-----G-----I-----F-----E-----V-C 490  
Db 5531 GSHLDSGPYRITAENELSGDTAIQIQRDPDRPPRPLIESIGTESLSLWKPAPWDGC 5590  
QY 491 -----W-----N-----A-----A-----G-----D----- 496

Db 5591 SDITNYVVERDDPLSSWIRVGNTRFTSMASVGLTPGKEYDFRIFADNVVYGRSDASPTST 5650  
QY 497 ---K-V-----G-A-----S-AS- 503  
Db 5651 LIKTESVKKKPIERKWEIDANGRLRGKADGPFVDYSDYVFDIYSKFPQPPVEISQOSV 5710  
QY 504 ---D---GS---V-C---VL---DL--RK 514  
Db 5711 YDRYDILEIGTGAGFVHRCRERSTGNIFAAKFIPIVSHSVKXDLIRRE 5759  
RESULT 8  
T15789  
hypothetical protein C41A3.1 - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C;Accession: T15789  
R;Bentley, D.  
submitted to the EMBL Data Library, November 1995  
A;Description: The sequence of C. elegans cosmid C41A3.  
A;Reference number: Z18404  
A;Accession: T15789  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-7829 <BEN>  
A;Cross-references: UNIPROT:Q18559; EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA8311  
C;Genetics:  
A;Gene: CESP.C41A3.1  
A;Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/3;  
2/2; 7514/3; 7596/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3  
C;Keywords: carrier protein; phosphopantetheine; phosphoprotein  
F;6547-6616/Domain: acyl carrier protein homology <ACP>  
F;2832,5271,6580/Binding site: phosphopantetheine (Ser) (covalent) #status predicted  
Query Match 70.5%; Score 2578.7; DB 2; Length 7829;  
Best Local Similarity 8.2%; Pred. No. 4.5e-52;  
Matches 456; Conservative 45; Mismatches 11; Indels 5083; Gaps 372;  
QY 1 MS---ISS-----D-----E-----V-N-F-----LV--- 12  
Db 1 MSACLVGSAKIPNGEDGHEMAENIFLNRNIALLPVKNYLSDFREKHEPVKAALVDGI 60  
QY 13 -----Y-----R-----Y- 15  
Db 61 EYFDDQYFGTGESEACMDPQQRMQLMGVYKGLNAGITLSEARVAVYTAACVYDK 120  
QY 16 --L--Q-----E-----SG--- 20  
Db 121 DLLPPDQYMATGNSASVMCGRITYFLNSRGAAGVIGTACSSSLVAFHLARQATQSGETKL 180  
QY 21 -----F---SH-----SA--F--TF----- 28  
Db 181 ALVCGANHVGRSFLSYNSHWSPNGRLAADRANGFVRAESFAVAVLCSKQFAENN 240  
QY 29 -----G-----I-K-S-----H----- 33  
Db 241 LLHCECVGSFAFNSDGTSLTAPNPIISOYEQLEAKNIDKDSVQLVTCGTGTGLGQ 300  
QY 34 -----I---S-----QS-----N----- 38  
Db 301 VELTAINRSFKDIRVMSPKSSMGHGAAGLIGVLSLYSMQHGIIPLNQLHLEPSED 360  
QY 39 -----IN-----G---ALV-P-P-----A--- 48  
Db 361 GEDKSMGFVNEEMELNRVAISSYGGGTNACAIIEKPEKPSLVQKESYAESNVLFSAKS 420  
QY 49 -----L-I-----S-----I-----I---QK--- 56  
Db 421 HESLKLQIEYTFQFMAQSDSAMEDIITYVNERKTYDFRAAVFGKNEBIAKRLQGDYS 480  
QY 57 ---IQ-----Y-----V-----EA---E----- 63  
Db 481 LTNLOESTPEVFEFGEGNEKLWLLRMLYEKNETFTHTVDKYCKLAETCGPFEARTALFFPF 540

QY 64 -----VS-I-----N----- 67  
Db 541 KLTPLTYNSRLISSMATFELLVQYNTLPNKRGLGQIFCLAVAKVITESAVQLI 600  
QY 68 -----E-----D-----G-----T----- 71  
Db 601 KGVBAENLTDLGLDIELSSKIPIBIOHLKSTKKILPIHISGELKETAKPNLMTFIVN 660  
QY 72 -----L-----FD-----GR-----P-----I 78  
Db 661 GBEILEDPVRKVOKLICQLFACGPDPAVKFRGRIUKTPYSPKQKQFPEVOTAMTWI 720  
QY 79 -----E-----SLS-----L-----I-----D-----AV----- 87  
Db 721 VDEQTNSSLSDAEISTVRTIVKQFLDIEDDINLLETGAVDLSLTSIEMVEAFGTAVNQT 780  
QY 88 MP-----:-----:-----D-----VV-----Q----- 93  
Db 781 MPFDLLEAYPTILNIVDFLKLTVTVPTVKATTSIHKKTSELSTSDINVIACDYQFAGVE 840  
QY 94 -----T-----R-----Q-----A----- 96  
Db 841 GEKELWDTLLTSLTKGKIDIRKKOCEGDAGLEVGLLKQDISMFDNSFFAIAXDAEFL 900  
QY 97 -----Q-----AY-----:-----:-----:----- 99  
Db 901 DPQHRLLNAAYNALKESGLTSIPDADLFLAISAHSEYRALAEKHINELDERLWMTVHS 960  
QY 100 -----R-----D-----KLA-----Q-----Q----- 106  
Db 961 MVAGRLAVLMGIRGRAMIVDTTCSSVATALEMAVKSIREGRFAIVATSLIQSSKWLVS 1020  
QY 107 -----Q-----:-----A-----A-----A-----A----- 111  
Db 1021 LKTLDDHSTNSFSDGSGFCRSDGVIIKTAEGDSAVIKISSAKSHHCGAVNTPV 1080  
QY 112 -----A-----A-----A-----A-----A-----AS----- 119  
Db 1081 SSISQLEAEAGSFSYVEGHGTATSDAGSAESMAYQKLGSELIMSSVKAQFGHCEVASGLI 1140  
QY 120 Q-----:-----Q-----Q-----G----- 122  
Db 1141 QLMKVSSIGKHGIPISVHNILPSEHINNIRLFPVAEEKQIDRSIAIVSFGITGKTIV 1200  
QY 123 -----SAK-----:-----:-----:----- 125  
Db 1201 VTERVSQLNVONIQCYLLPVSAKTKOGLKACLSLIEMIDNSCESLYDISTTLQOKT 1260  
QY 126 N-----G-----E-NT--AN-----G-----E----- 134  
Db 1261 NFKWRTAVGSSHADVVVLKQKFLTSEHTSLTNWHISTSHSISGCTFFHNIPEFEDHY 1320  
QY 135 -----E-----NG-----A-----H-----T-----IA-----N----- 143  
Db 1321 SMFCHRLRPEPHSNTNSIVHLLAVVALIRVILKHLKLTNSFAVGPNLSLIVLAAVDAAPS 1380  
QY 144 ---N---HT---D---MM-----E-----V----- 151  
Db 1381 HYLNDLLHAFANDVVRMKRIARDVTISENVKLNLNAGEPITTARQAVEATIDOKVKV 1440  
QY 152 -----G-----D-----G-----D-----V-----E----- 156  
Db 1441 RLPETTLILSPRAYEFASOLETIQDYKYLIGEKPQGGQVDFAGIFGTPKILDLPEYP 1500  
QY 157 -----IP---P---N-----KA-----VVL-----R- 166  
Db 1501 FNRKSWLPIDPSVPSNEKGEKPLPKSYEFLLKSQKHQVHVVDISKVLPGATSIRL 1560  
QY 167 -----:-----:-----:-----G----- 167  
Db 1561 VHOLNGKPTVELSNIDFLNKTIPSEAPSVVKIEBQDGLKLVFGETDAISPKLTQLQNFN 1620

QY 168 -----H-----E-----S-----EV-F----- 173  
Db 1621 PIPNERLNAEVHHTDNIYERFANSHLTYRNEFQMVDSLKYTMGKGVEFRFVSMKDLID 1680  
QY 174 -----I---C-----:-----:-----:----- 175  
Db 1681 GTLQAIVGCFYFFENTNDNSFPFTIDQLSILNGDISQQLHAVALKYDSSGNFINGDATV 1740  
QY 176 -----:-----:-----A----- 176  
Db 1741 YDALGNIIILHISNVTFKRLNGQSAPLSTKTVDSKTKKVENEDQKRASKMHLHWFE 1800  
QY 177 -----W-----N-P-----V-----S-DLASS----- 187  
Db 1801 NFGWTIDIDNTTGFDDLGLTSIQAVKLRNAIKSNYPNASSTCVDPYPSIDLL-SCYLSTLN 1859  
QY 188 -----S-G-----D-S-----T-----A-R-----I-----W----- 196  
Db 1860 DPQVETSTGDDIQKOLTEDHKPRLAENPIGVMAAACRLPGGVSSPSSELWELLKIGKN 1919  
QY 197 -----N-L-----S-----:-----:-----:----- 199  
Db 1920 ASSRIPATVPTRNTLISGSKYGNPVEGCFITQDVTQFDPSPFFKISKSEAEILDPQORL 1979  
QY 200 -----ENS-----TS-----:-----:-----GS----- 206  
Db 1980 LLECVQECLENGSVIETSNVGVFGLMEKXYQDMMESSILAMLGSMAAVAGRVNVI 2039  
QY 207 -----:-----:-----:-----T-----Q----- 208  
Db 2040 CYGPSVTIDTACSSSLVALEMAINALLDNRCISKVIVAGVNLILNEKGOGLRTNGKLSQH 2099  
QY 209 ---L-----VL-----:-----:-----R----- 212  
Db 2100 GMSLSDFRASGYGRSDGCVLMLLELAKENFHYMSTIQSVNVNHGGRSVSLTAPNGVAHK 2159  
QY 213 -----:-----:-----:-----H----- 213  
Db 2160 MLLTSVINGQPSLAIDYWEAHGTGTPGDPPIBFTLSSILQNIIGSVKASLGHGEASAG 2219  
QY 214 -C-----IR-----E-----G-----G----- 219  
Db 2220 TCGLLKLFLMLTYQVPTLIHFHVLNKOINAGSIRLPIIGEDSELVSAGISSFGVSGTNA 2279  
QY 220 -----Q-----DVP-----:-----:----- 223  
Db 2280 AAIAFNDNNKLEPYPIHKYIYILPISAKNQISLDNLEKQILSVIPLTDVPCINIASALAN 2339  
QY 224 -----SN-----:-----:-----KD----- 227  
Db 2340 NRSHFTIRNALIVNSGIVNSKMEGKPHRVAKKDRVHVVKLDCSLDASLLQYDVINETY 2399  
QY 228 V-----TSL-----D-----W-----N-S-E----- 236  
Db 2400 VASLKNQSFAMKFAIIRKFLTSLSLEYIEIVASDGBELLAVLLANGSLKWFENFKTMIE 2459  
QY 237 -CTLA---A-----:-----:-----TG-----S 244  
Db 2460 IGSLLTEFADHLNSTSSSIKSYQTPESHNDSPMELMKLIMKLYITGYVDVWATVYS 2519  
QY 245 -----Y-----D-----G-----F----- 248  
Db 2520 PVEQFIALPNYQFNKOTLWFEERLEIVDHYLIGTIDEBSEDTLILKNQISELRHPQFKG 2579  
QY 249 -----A-R---I-----W-----T----- 253  
Db 2580 KPLDVGTWSEIAIBALKIRNEIPFSIQNLKLTBLITLTKPAWLTNVTNRNEDDEGPNVSA 2639  
QY 254 -----:-----:-----K----- 254  
Db 2640 IDQRLFSLNASSVEIQNIEVPAVEVQIPDKVVYLKECPNAVIRHRNMVYVDSRAEQSP 2699  
QY 255 -----:-----:-----DG-----N----- 257

Db 2700 FRTANIVLEIIGFAPTPSDMFEILGLVPSVHMVQVDDGALWQFQWISQDKXVLSNIY 2759  
QY 258 -----LG-----D-----R-----P-----I-----KT-----F- 345  
Db 2760 VLKDAKGLEIPTIRMHKKSTLLSSQEASIVAAKTQMAVRHKVCLAVGDVIESGLDIDES 2819  
QY 260 --ST--LG-----HTN--E-----VN--A 354  
Db 2820 QLSTGFSELGDSLATVDLLNRLNQYPPPIELTTSDLPNPSIIDLSIMEQLLNEKGI 2879  
QY 264 -----K-----W-D-----PT----- 360  
Db 2880 TEPSEPNTKSLRGRKLSIPAVRAQVLAQIEFVENVNSKQKEVQAEAPSSSECSNMLE 2939  
QY 266 -----GNLL-A-----S-----C-----S----- 368  
Db 2940 ESDATVDRTEIRRKVSLAVFDLATETLSAEDLOSQKFTGLGMDLSIVDFNRLNDKYFP 2999  
QY 269 -----IF-----A----- 271  
Db 3000 DDEITASDIFDPTVDELSDHIVRKKSVPDPAASEIMKETMNGISTSDAETHLENLS 3059  
QY 272 -----LK--W--N--K--K--GNFI--LSA-----G----- 285  
Db 3060 QSFMLLENQNSINPTLKMIWSNOTIKLVKPSDGNFLFELNANGQKEIKHFTGPNII 3119  
QY 286 VD-K-----T-T-----I----- 291  
Db 3120 IDLKGHEGSTETLYMSLLNLVKSISKLEIQCRFGVQSEFGLGNSISRAFMTVAAEKNP 3179  
QY 292 -I--W-----D----- 294  
Db 3180 LISFAMVQNVQSVFVDSDSPTGNMLITGGLSGIGLEIGKFIANNGAENVILISRRQPT 3239  
QY 295 --A-----HTG----- 298  
Db 3240 AKALRDILSTELTHIGLAKTIVLKLKNISAKLIVQSPKLSFTSKYISDLPSKKVTF 3299  
QY 299 -----E----- 299  
Db 3300 FYNLOSCKPSKHFHMLFEWLEVYLKQMTVHTTAADINDKEKLIRELTKLVNGITGIH 3359  
QY 300 -A-----K--O--Q--P--PH----- 307  
Db 3360 SAGVLKDKLIERQKESFNQVFTPKANGFHVLEIEKHFNYKIENFIMMSFTAACNEG 3419  
QY 308 -----S----- 308  
Db 3420 QLVNGVSNAYLEYQVORRRQKSGCAIQWGNWIDTGMATDENVRKFLANLGLGQHND 3479  
QY 309 -----A-----PAL-----DVM-----Q-----S 318  
Db 3480 ALKYLACILTKPELIMVANIDWNVLKNRKLDPKOLINTGILPFEDFTGKNNESEFPPLS 3539  
QY 319 N----- 319  
Db 3540 NGDFEKMNFVSEDEEVELELIKVKVSSILMCSPTKLKNNKIMDMGLDSKLIIVBFLNF 3599  
QY 320 -N-TF-----A-----S-----C----- 325  
Db 3600 INSTFKISVNLSDAYNHPTLEKLAHI FEQMTIVDHPVNSKVEEIPKSTDFCFIFGINI 3659  
QY 326 -----ST-----D----- 328  
Db 3660 FFDKNDFDRAKSTAVKLENGEQLPTAGYAVSVVGSIRDVVSKIEAAPQOIKLCQE 3719  
QY 329 -----M-----C-----I----- 331  
Db 3720 SSKCVLMLTCQSQYPMNGQLVENYEI FRTTLOSLCKKCDXYLQGDVSLWEILFNTHY 3779  
QY 332 -----H--V--C----- 334

Db 3780 KLLQLTKHQPIMFCFYATAQLWSLIGIVPDYVLGHVSGELVAGVLGIMSTEDGLRLI 3839  
QY 335 -----K-----LG-----Q-----D-----R-----P-----I-----KT-----F- 345  
Db 3840 VERGKAMENIAGLALLAQVREIADEVLRKFVKVSVATINSKPQVFBAGTKSVLDAALAFV 3899  
QY 346 --QG-----HTN--E-----VN--A 354  
Db 3900 KGQKQATVYNQOYPFHSLNIQETHLVLSRQCLADIKFSAGRTPLVSNVTGQIINTFSEA 3959  
QY 355 -I-----K-----W-D-----PT----- 360  
Db 3960 YIVKHTVSAVKFVDCVETLQAKGVTWIDAGSAAVLATFVKRIIQPTLSKHRIVQTCKE 4019  
QY 361 -----GNLL-A-----S-----C-----S----- 368  
Db 4020 KESDVNQLVQACLEBQSLPISWTTLYGCGRNADERLVEFPMTHNDIINKNDEFELLEGH 4079  
QY 369 -----DMT-- 372  
Db 4080 QLVNGKIIVAGAYQLFKIDQLVKLKAAGMELMLKNVKFLKPWYIEDNREYQIOWNSDWTIE 4139  
QY 373 -----LK----- 374  
Db 4140 LIVNSVIVCSLEVEPQNSVLKLETISENEKPEVHDFYETLFRNGLYQDSGFRRIESARR 4199  
QY 375 -----I--W-----S--MK----- 379  
Db 4200 SDKRCFSQIKSSPFWPLIDSAMHSITASVVRPRPCYFLPVAMGVTWKTDSFTLPLNL 4259  
QY 380 --Q-----D--N--C-----VH----- 385  
Db 4260 HAGTVITSETDKFIQVNVALLAGDTPICEVRNMTIVVLKTEPVHTRIPNSIETVETPPKS 4319  
QY 386 -----D--L-----Q-----Q--HN--K----- 392  
Db 4320 EIEIVGFDISLPYNQISENSENWOHLKTNTVKQLHNRSLKQDHARVALLDSARYDPE 4379  
QY 393 -----E-----I--YT-----I-----K 398  
Db 4380 YGIRPSEAKFIDPOORLLCSVAKLLDSLLITSLTNTSGVFIGCSANFESHVYAYGK 4439  
QY 399 -----WS-----P-----TG----- 403  
Db 4440 DPRAEWSGTSNSALAGRIAHWLKLGPPVTLDTACSSSFYALSAAADALRTQCEYAV 4499  
QY 404 -----P----- 404  
Db 4500 GTVNLVHMTTDLVQLNAKMTVDDFCKAFDNDANGYKRSSEAVGCSMLLTSPNIDSVATIT 4559  
QY 405 -----GT-----N-----NP-----NA----- 411  
Db 4560 NYATGHNGTSSSLFTPNGLSQLEVMOARATNPLEKILEIOTHCTGTGLGDPIDIEINASKLV 4619  
QY 412 -----N-----LM----- 414  
Db 4620 SSACKIGSVKSNIGHTEGSSGLVSLCSSLMSFRSKYRVAQLHLKCTNSIKTNKMICRFI 4679  
QY 415 -----L-----A-----SA--S----- 419  
Db 4680 GEDADENNSILINNFQGTGNSCVVLKPKNAISEHFVSSEVFPYILLSSHSKSLQKYVQ 4739  
QY 420 -----F-----D--S-----T-----VR----- 425  
Db 4740 VLCEPISNSAKSLHDIMMSLFQKKIHVHQFIIFNPKRIAVTSLDGFVEVRDERLEKLK 4799  
QY 426 -----L-W-----D-VDR-----GI----- 433  
Db 4800 HPCSFLKEGVVHFDKXSFQVLDLPISIVFNNTLHWALDSYRDEIDHESQMSFKNIFY 4859  
QY 434 -----CI-----H----- 436  
Db 4860 EKVLETPMPQNPISKVVICRGLDLIPKIEIDEVSSFANGIIVFPHIPNSIPEYLKLA 4919

QY 437 --TL-----TKH-----Q-----E-----P-----V----- 445  
Db 4920 VMSLISRNQNVFIICCFENGTSHTWTGTLRLSLASEKMFYKVFSDIKVDALKLFENHE 4979  
QY 446 -----YS-----V-AP-----S-----P----- 452  
Db 4980 YMFEAIFYKRGYRVERLRKVNPTFKPAQYKCLISGGTGGIGSAIINELPKSSVIIT 5039  
QY 453 -----DG-----R-----Y-----LASG-----SFD----- 463  
Db 5040 RKNIASEDGKTFLLSSDITRLDISHKFNYPHILA-GIVNNSLHENVKRDSLDENVSKLOG 5098  
QY 464 -----KC----- 467  
Db 5099 AKNLMKCCDTSHFVSSSIANVLGSYGQSNYAFSNGLVTSFLETSTKSTIIHWGPWKD 5158  
QY 468 -----I-----W-----N-----TQ-----T-----G-----ALV-----H 478  
Db 5159 VGMLAQPERRERIVKQIESNGWKLPPNQDAISVFYTFQMETHEQIIIVFDGDFDTIVARQPH 5218  
QY 479 -----S-----Y----- 478  
Db 5219 LOKLSEVVEKTKVKEIKKKSLNFEIPEIIVGIIITDSSKLNIPFMDLGDLSLCLMENLR 5278  
QY 479 -----S-----Y----- 481  
Db 5279 YSLNKNPDLLELVSEMFENATYKLTQYVETLRLKAKHSLSLSDHRYSSQVSNKEDTRVA 5338  
QY 482 -----G-----TG----- 484  
Db 5339 VIGWSAEFGSSNIHEYENLMDGICSTGNKYLKPNFQDNKFNLTDEADARVLDPOV 5398  
QY 485 -----CIF-----E-----V----- 489  
Db 5399 RXPQHAYLALENSGVYKQKHELRCGVFAGAEPSDYGRADDDHDAWKLFVMMNMSYLAS 5458  
QY 490 -----C-----W-----NA-----AG-----D-----KV-----CA----- 500  
Db 5459 YASYCLDLKGEAVSVYSACSTALVANAVAKSIQSGSMYALVGAASIAEVSGALSGFDD 5518  
QY 501 -----SAS-----D-----GS-----V-----C-----VL-----DL----- 512  
Db 5519 QKKTMPKSGVCRPFDPKDEGIVRGSGVGVFLKRYSQALLNDNDNVHVFVIFKDFAINNDGH 5578  
QY 513 -R-----K 514  
Db 5579 SRASFAPNPAGQLK 5593

RESULT 9  
T42737  
gp330 protein precursor - rat  
N:Alternate names: megalin  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
R:Accession: T42737  
R:Saico, A.; Pietromonaco, S.; Loo, A.K.C.; Farquhar, M.G.  
Proc. Natl. Acad. Sci. U.S.A. 91, 9725-9729, 1994  
A:Title: Complete cloning and sequencing of rat gp330/megalin, a distinctive member of the gp330 family  
A:Reference number: A58173; MUID:95024033; PMID:7937880  
A:Accession: T42737  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4660 <SAI>  
A:Cross-references: UNIPROT:P98158; EMBL:L34049; NID:g561852; PID:g561853; PIDN:AAA51369  
A:Experimental source: strain Sprague-Dawley; kidney  
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding  
F:1-25/Domain: signal sequence #status predicted <SIG>  
F:46-4660/Product: gp330 protein #status predicted <NAT>

Query Match 70.4%; Score 2574.9; DB 2; length 4660;  
Best Local Similarity 9.8%; Pred. No. 2e-52;

Matches 426; Conservative 66; Mismatches 14; Indels 3841; Gaps 368;  
QY 1 M-S-----I-----S-----SDE-----V----- 8  
Db 113 MTCNSGQCIPSEYRCDHVSDFCDPDSDERNCHYPTCDQLTCANGACACNTSQRCDQKVDGRD 172  
QY 9 -----N-----F-----L-VY-----R-----Y----- 15  
Db 173 SSDEANTTLCQKEPECGSGECILRAYVCDHDNDCEDESDNRCNVCYDTGCGHQFTCSNG 232  
QY 16 --L-----Q-----ES----- 19  
Db 233 QCINQNWVCGDDDCQDSGDEDCESNQSHHRCYPREWACPGSGRCISIDKVCDCGVDPDCP 292  
QY 20 -----G-----F-----S----- 22  
Db 293 EGDENNVTSGRTCGMGVCSVLNCEYQCHQTFPGCEFCPPGHIINSNDSRTCIDFDDCQ 352  
QY 23 -----H-----S-AFT-----F-----G----- 29  
Db 353 IWGICDQKCNQGRHQCCEGYILLERGOHCKSSDSFSAASVIFSNGRDLLVGDHLGRN 412  
QY 30 --I-----KS-----H-----I-----SQ-----S-----N----- 38  
Db 413 FRILAESKNRGMVGVDFHYQKRVFWTDPMQEKVFSTDLINGLNTQELILNVSVDTPENLA 472  
QY 39 --IN----- 44  
Db 473 VDWINKLYLVETKNRIVDVNLEGNQVRVLTITENLGHPRGIALDPTVGYLFFSDMGSL 532  
QY 45 --P-----PA-----ALIS-----I-----IQ-----K----- 55  
Db 533 GQPKVERAFMDGSRNKKDLVTKVGHGAGITLDLVSRYVWDSRYDIETVTVDGIOKRT 592  
QY 56 -----G-----L----- 59  
Db 593 VARGSLVPHFPFGISLFEHVFFTDWTQAMVMAKSKFTETNPQVYHQSLLRPHGVTVYHA 652  
QY 60 -----V-----E-----V-----AE-----VSI-----N----- 67  
Db 653 LRQPNATNPGSGNNGCAQCVLSHRTDNGGLGYRCKCEFGFELDDDEHRCVAVKMLLF 712  
QY 68 -----E----- 71  
Db 713 SKTAVRGIPFTLSIQEDVMVPVTSPPSFFVGIDFDAQHSTVFYSDLSKDIIYKQIDGT 772  
QY 72 -----L-----FD----- 74  
Db 773 GKEVITANRLSEVECLTTFDWISRNLYWTDDGLKSVTLRLADKSRRIISNLNPRISVV 832  
QY 75 -----G-----RP-----I-----E-----S-----L-----I-----DA----- 86  
Db 833 HPTAGYMFUSDWFRPAKIMRAWSDGSHLMPIVNTSLGWPNGLAIDWSASRLYWDVAFDK 892  
QY 87 -----V-----M-----P-----D-----V-----VQ-----T-----R----- 95  
Db 893 IEHSTLDGLDKRLGHVDQTHPPFGLTVFKDNVFTDWRGALIRVKSDDGDMTVIRRG 952  
QY 96 -----Q-----Q-----A-----Y-----RKDL----- 103  
Db 953 ISSVMHVKAYDADLQTSNYCSQTTTHANGDCSHFCFPVPNFQVRCVCPYGMKLRQDMTC 1012  
QY 104 -----A-----QO----- 109  
Db 1013 EGDPAEPPTQCGSLSPFNNKGVPSFFRCGVDDCHDNDSEHOCGVNNTCSAFA 1072  
QY 110 -----A----- 110  
Db 1073 CVRGQCIPGQWHCDRQNDCLDSDGDEQNCPTHTATSTCTPSTFTCDNHVCIPKDWCDTD 1132  
QY 111 -----A----- 116  
Db 1133 NDCSDGSDENKQASGTCQPTQFRCPDHRCISPLYVCDGDKDCADGSDGAGVNLNCTSAQ 1192

QY 117 ---A--AS-----Q-Q----- 121  
Db 1193 FKADGSSCINSRYRCGVDCRDSBAGCPTRPPGKCHPDEFQCGDGTCPINTWECD 1252  
QY 122 ---GS---A-K-----NG-----E----- 128  
Db 1253 GHPDCIHGDEHTGCVKPTCSPTFLCDNGNCIYKAMICDGDNDCRDMSDEKDCPTOPFH 1312  
QY 129 ---N-TA-----NG--E--E-----NG--AH----- 139  
Db 1313 CPSTQWQPGYSTCINLSALCDGVDFCPNGTDESPLCNQDSCSHFNGGCTHQCWQGPFGA 1372  
QY 140 ---T-----I-----A----- 142  
Db 1373 TLCPLGYQLANDTKCEDINECDIPGFCQHCVMNRGSPRCACDPBYTLESGRCTKV 1432  
QY 143 ---N-----N--HT-----DMW----- 149  
Db 1433 GSENPJLVASRKIIVDNITAHNLSLVQDVSFVVALDFDSVTCGRVFSWDLLOQKTW 1492  
QY 150 ---E--VD-G--D-----VE-----I----- 157  
Db 1493 SVFQNGTKRVVHDSGLSVTEIADWIGRNLWYTDYALETIEVSKIDGSHRTVLISKV 1552  
QY 158 ---P--N-----KA-----VV-----L-----R-- 166  
Db 1553 TKRGLALDPRMGDNVWFSDWGHHPRIERASMDGTWRTVIVQEKIYWPCLSIDYPNRL 1612  
QY 167 ---GH--E--SE-V-----F----- 173  
Db 1613 IYFMDAYLDVIEFCDYDGHNRQVIAISDLVHLHPHALTLFEDFYWTDRTQVMOQANKW 1672  
QY 174 ---I-----CA-----W 177  
Db 1673 HGGNSVVMYSVHQPLGITAHPSPRPPSPRNPASCSHCLLSAQAPRHSACAPSGW 1732  
QY 178 N-----PVS-----D----- 182  
Db 1733 NLSDSVNCVRGQDFLMSVRDNIIFGISLDPEVKNDAWMPISGIGHQYDFDSEQF 1792  
QY 183 -L-----LA--S--G--D--S-----T-A----- 193  
Db 1793 IYVWNPGEIHRVKTGDSNRTPAPLISGLSLGLALDWVSRNIYVTPASRSIEVLTK 1852  
QY 194 ---R-----I-----W----- 196  
Db 1853 GDRYKGLIANDGTPLGVPVGVIAVDPAARGKLYMSDHDGTSVPKATIASANMDGTSK 1912  
QY 197 ---NL-----S--E--N--ST-----S-----GS 206  
Db 1913 ILFTGNLOHLEVTLDIQEQKLYWAVTSRGVIERGNVDGTERMLVHHLAHPWGLVYGS 1972  
QY 207 ---T--QL-----VLR-----H-----C-- 214  
Db 1973 FLYVSDQYEVIERVDKSGNKKVLRDNVPLRGLRVYHRRNAADSSNGCSNPNACQ 2032  
QY 215 I-----R-----E--G----- 218  
Db 2033 ICLFVPGMFSCACAGFKLSPDGRCSPPNSFMVMSLPAVRGFSLELSDHSEAMVPVA 2092  
QY 219 GQ-----DV-----P--SN-----K-- 226  
Db 2093 GQGNVLHADVDVANGFIWCDPSSSVRSNGIRIKPDGNSFTNVVYIGANGIRGVA 2152  
QY 227 -D-----V--TS-----L-----D-----W----- 233  
Db 2153 LDWAAGNLFTNAPVYETLEVLRIINTYRVLVKVSDMPRHIIYDPKHYLFWADYG 2212  
QY 234 -----N-----SEG--T--LLA-----TGSY-----DG----- 247  
Db 2213 KPKIERSFLDCTNRVLVSEGIPTPRL-AMHDHTG-YIYVWDDSLDLIARIHLDCGESQ 2270

QY 248 ---F-AR-----IWTKD-----GN----- 257  
Db 2271 VRYGSRPYPTGYGITVFGESIHW--DRNLKKVFOASKQPGNTDPPVVIRDKINLLRDVTI 2329  
QY 258 ---L--A-----S-----TLG-----Q-- 264  
Db 2330 FDEHAQPLSPAELNNPCLQSGGSHFCFALPELTPRCGCAFGTLGNDGKSCATSOED 2389  
QY 265 -----H-----K----- 266  
Db 2390 FLYLSNLSRLSLFDRDHSLPFQVIVAGTAIALDYDRNNRIFFTOKLSLRGQISY 2449  
QY 267 ---G--P-----I--F----- 270  
Db 2450 VLSYSSSPTVLLSNIGVTGDIAPWINRRIYYSDFSNQTSMAEDSGSNRAVIARVSK 2509  
QY 271 --A-----LK--W--N--K--K--GNF--IL--S--AG-----V 286  
Db 2510 PRAVLDPCKGYMYWTDWGTNAKIERATLGGNFRVPIVNTSLVWPNGLALDLETLLYWA 2569  
QY 287 D---K-----T-----TI--I--W-----D--A 295  
Db 2570 DASLOKIERSTLTGTNRVVSFAHSGFLTIVYQYIYTDLYTRKIYRANKYDGSDLVA 2629  
QY 296 -----TG-----E--A-- 300  
Db 2630 MYTRLPTQSGISTVVKTORQOCNPDQFNGGCSHICAPGNGAECQCPHEGNWYLAND 2689  
QY 301 -K-----Q-QF-----P-----PH--S--A-- 309  
Db 2690 NKYCVVDTCRCNQLOFTCLNGHCINQDWKCDNDGCGSDELPTVCFAHTCRSTAFTC 2749  
QY 310 ---P-----A-----L----- 312  
Db 2750 GNGRCVPYHYRCDYNDGNSDEAGCLFRNCNSTTEFTCSNGRCIPLSVVCNGINNCHD 2809  
QY 313 -----D-----V-----D----- 315  
Db 2810 NDTSDKNCPPHTCPDFTKCTQTNICVPRAFLCDGNDGDSDENPIYCASHTCSNE 2869  
QY 316 -----W-----Q-----S----- 318  
Db 2870 FQCLSPQRCPISYWFCDGADGADGDEPDTCCHSVNTCRASQFQCDNGRCISGNWVCDG 2929  
QY 319 -N-----N--T--F----- 322  
Db 2930 DNDGDMDEDRHCELQNCSTOPTCVNRPNNRCIPQYVWVGDDADCDALDELON 2989  
QY 323 -----A--SC----- 325  
Db 2990 CTWRTCSAGEFSCANGRCVRSQSPRCDRRNDGDSYSDERGCSYPPCHANQFTCONGRICIPR 3049  
QY 326 -----S-----T-----D--MCI-----HV--C-----K 335  
Db 3050 FFVCDENDGCGSDBEQHEHLCHTPTCPLHQFRCNDGNCIEMGRVCNHNVDGSDNSDEK 3109  
QY 336 -----L--G-----O-- 338  
Db 3110 GCGINECLDSSISRCDHNCTDTITSPYCSCLPGYKLMDSKRSVDIDECKESPQLCSQKC 3169  
QY 339 -----D--R-----P--I----- 342  
Db 3170 ENVVGSYICKCAPGYIREPDGKSCRONSNIEPYLIFSRYRINLTDDGSSYSLLIOGLG 3229  
QY 343 -----KT----- 344  
Db 3230 NVVALDFDRVEKRLYWADEAKOIIERMFLNKTRETIINHRLRAESLAVDWVSRKLYWL 3289  
QY 345 -----F-----O--GH----- 348  
Db 3290 DAILDCLFVSDLEGRHRKMAIOACHVDANNTFCFHEPRGIVLHPQGHVWADGWVHAYIG 3349  
QY 349 -----TNE--V-----NAI-----KW-----D----- 358



Db 3350 RIGMDGTNKSIIISTKIEFWNAITIDYTNLLYWAADAHGLYIEFSDLEGGHRRHTVYDGLS 3409  
QY 359 --P--T--GN--L--LAS- 366  
Db 3410 PHFPALTIPEDTVFTWTDWNRTRTEKGNKYDGSRRVVLVNTTHKPPDIHVHPYRQPMGN 3469  
QY 367 ---CS---DD---M---T---L---K--- 374  
Db 3470 PCGTNNGGCSHLCLIKAGRGFTACPDFFQTVQLRDRTLCPMPCSSSTQFLCGNNEKCP 3529  
QY 375 IW---S---MK---Q---D---NC---V 384  
Db 3530 IHWKCDGQKDCSGSDPEPLCPHFRCLGFCQCRDGNCTSPQALCNARQCADGSDSDRV 3589  
QY 385 ---H---D---L---Q---Q---H---N--- 391  
Db 3590 LCEHRCESNEWQCANRCIPQSWQDCSVNDCLDNDEDTSHCASCRTCPGQKCNNGRC 3649  
QY 392 ---K---E---I---Y---T---IK-W--- 399  
Db 3650 IPQSWKCDVNDGDSDEPIDECTTAAYNCNDNHTEFSCKTNVRCIPQWAVCGNPGDCRD 3709  
QY 400 ---S---PTG---P---GT---N---N--- 408  
Db 3710 NSDEQCESVPCHPSGDFRCANHHCIPLRWKCDGTDGCGNDSDEENCVPRECSSEFRCA 3769  
QY 409 ---P---N---AN---LM---LA---SAS 419  
Db 3770 DQOCISRWVQDQENDCGNDSBERDCMTCHPEHFQCTSGHCVCPKALACDGRADCLDAS 3829  
QY 420 FD-S---TVR---LW---D---VDRG---I---C---I--- 435  
Db 3830 -DESACT-RFPNGTYCPAAMFEKNHVCIQSPWICDGENDCVD-GSDEIHLCPNICE 3886  
QY 436 ---HTL---T---K---H---QEPV---Y---S--- 447  
Db 3887 SPORFCDNSRCVYGHQLCNGVDDCGDGSDEEHCERKPTCTDTEYKCSNGNCISQH 3946  
QY 448 ---V---A---A---P---S---P--- 452  
Db 3947 YVCNVDNCGDLSDETGNLGNDRNCAENICEONCTQLSSGGFCISCRFGFKPSTSDKNS 4006  
QY 453 --D---G---RYLAS-GS--F--D--K---C--- 466  
Db 4007 CQDINECEBFGICPQSCR-NSKGSYECFCVDGFKSMTHYGERCAADGSPPLLLPENV 4064  
QY 467 ---HI---WN---T---Q---T---GA--- 475  
Db 4065 RIRKNTSSEKFESELEEHEIQTIDYDWDPEHIGLSVVVYTVLAQSGQFGAIKRAYIPN 4124  
QY 476 ---L---V---H---S---YR--- 481  
Db 4125 FESGSNNPIREVDLGLKYLQMPDGLAVWVGRHIYMSDAKSQRIEATLDGRYRKWLIT 4184  
QY 482 ---G---T---G--- 484  
Db 4185 QLDQPAIAVNPGLMFWTDQKQPKIESAWNGEHSVLYSENIGWPNGLSIDYLND 4244  
QY 485 ---G---IFE---V---C-W--NAAG- 495  
Db 4245 RVYWSKEDVIEAIKYDGTDRRLIINEAMKPPSLDIFEDKLYWAKEKEVWRQNKFGK 4304  
QY 496 ---DKV---GA-S-A---SD---G 505  
Db 4305 ENKEKVLVNPMLTVQRIHFQLRYNOSVSNPCKQVCSHLCLLRPGGYSACPGSGDFVTG 4364  
QY 506 S-V-C---V---LD--- 511  
Db 4365 STVQCDAASELPTMPPPCRMHGGNCYFDENELPKCKSSGSGYCEVGLSRGIPPGT 4424  
QY 512 -----LRK 514

Db 4425 TMAVLLTFVIVIVGALVLVGLFHYRK 4451  
RESULT 10  
A88852  
protein unc-22 [imported] - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: A88852  
R:anonymous, The C. elegans Sequencing Consortium.  
Science 282, 2012-2018, 1998  
A:title: Genome sequence of the nematode C. elegans: a platform for investigating biology  
A:Reference number: A75000; PMID:99069613; PMID:9851916  
A>Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C.elegans  
A>Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and  
A:Accession: A88852  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-6831 <STO>  
A:Cross-references: UNIPROT:Q23550; GB:chr\_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN000  
C:Genetics:  
A:Gene: unc-22  
A:Map position: 4  
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology; I  
Query Match 70.3%; Score 2572.2; DB 2; Length 6831;  
Best Local Similarity 8.0%; Pred. No. 5.1e-52;  
Matches 472; Conservative 29; Mismatches 10; Indels 5358; Gaps 394;  
QY 1 M-----S-----I-----S-S-----D----- 6  
Db 784 MTEAERRQSLFPQKVKWKMDIPLPEKTQQQVDKICEWCKTYSRPNKIRWYKDKKEIFS 843  
QY 7 -----EV-----N-----FL----- 11  
Db 844 GGLKYKIVIEKNVCTLIINNPEVDDTGKYTCBANGVTHAQLTVLEPPMKYSFLNPLNPT 903  
QY 12 --VYR-----Y-----L----- 16  
Db 904 QBIYRTQAVLTCKNTTPRAPLVWYRGSKAIQEGDPRFIEKXDAVGRCTLTIKEVEEDQ 963  
QY 17 -----Q-----ES----- 19  
Db 964 AEWTAIRITQDVFQKQVYVEEPRHTFVVPKMSQKNESDLATLETVDNDKDAEVMWHDG 1023  
QY 20 -----G-F--S-----H-----S-----A-F----- 26  
Db 1024 KRIDIGVKFKVSSNRKRLLIINGARIEDHGEYKCTTKDDRTMAQLIVDAKNKFIVALK 1083  
QY 27 -----T---F-G--I---K---S---H----- 33  
Db 1084 DTEVIEKDDVLMCQTKDTKTFGIWFRNGCKQISSMPGGKFETQSRNGTHTLKIGKIEMNE 1143  
QY 34 -----ISO-----S-----N----- 38  
Db 1144 ADVYEIDQAGLRGSCNVTVLEAKRPILNWKPKIEAKAGEPCVVKVFPQIKGTRRGDPK 1203  
QY 39 --I--NG-----AL----- 43  
Db 1204 AQLKNGKPIDEMKRLVEVIIKDDVAEIVFKNPQLADTGKWALELGNAGTALAPFEL 1263  
QY 44 V-----PP-----AA--LI--S-----II--Q-----KG 56  
Db 1264 VKDKPFPKGPPLTKNVTAEGLDLVWGTPDDEGAPVKAIIEMQGRSGNWKVGETKG 1323  
QY 57 -----L---QY---VEA-----E-V-----SI-----N----- 67  
Db 1324 TDFKVKDLKEHGEYKFRVKALNECGLSDPLTGESVLAKNPYGVPGPKNMDAIDVDKHC 1383  
QY 68 -----EDG-----T-----L----- 72  
Db 1384 TLAWPEEDGGAPITGYIIERREKSEKDWQGVQTKPCCCLTDKVVVEDKEYLYRVA 1443

QY 73 -----F-----D----- 74  
Db 1444 VNKAGPDCDHGKPIKWKAKKASPEFTGGGIGKDLRLKVGETIKYDVPISGEPLPECLWV 1503  
QY 75 ---GRP-----IE-----S-----L-----S----- 82  
Db 1504 VNGKPLKAVGRVKMSSERGHIMKIENAVRADSGKFTITLKNSSGSCDSTATVTVGRPT 1563  
QY 83 ---L-I---D-AVM---PD---VV-Q-----T- 94  
Db 1564 PPKGFLDIADVCAGATLSMNPDDGGDLPTGIVIEAQDMNDKNGKIVIEVKVDPNTTIL 1623  
QY 95 -----R-----Q-O----- 97  
Db 1624 KVLGRKNGVKFRVAVNNEGESEPLSADQYTOIKDPWDEPGKGRPEITDPPADRIDI 1683  
QY 98 -----A---Y---RD----- 101  
Db 1684 AWEPPHKDGGAPIEBEYIEVRDPTDKWEKVKRVPDNTNAGISGLKEGKEYQFRVAVNKA 1743  
QY 102 -----K-LA----- 105  
Db 1744 GPGQSPSEKQLAKPKFIPAWLKHDLNLSITVAGATVRWEVKIGGPIPEVKWFKGNQ 1803  
QY 106 Q-----Q-----A-----A-----A----- 110  
Db 1804 QLENGIQLTIDTRKNEHTILCIPAMRSDVGEYRLTVKNSHGADDEKXANLTVLDRPSKN 1863  
QY 111 -----A-----A-----A-----A----- 113  
Db 1864 GPLEVSDVFEDNLNLSWKPPDDGGGEPIEYVEVEKLDATGTRWVPCAKVKDKTAHIDGLK 1923  
QY 114 -----A-----A-----A-----A----- 118  
Db 1924 KGQTYQFRVAVNKEGASDALSTDKTKAKNPYDEPGKTGTTPDVVDADRVSLWEPPK 1983  
QY 119 S-----Q-----Q-G-S-----A-----K-N-----GE 128  
Db 1984 SDGAPITQYVIEKKGKHGRDQECGVSGDQTNABEILGLKEGEYQFRVAVNKAAGPGE 2043  
QY 129 -----N-----T-----AN-----GE-----E-----N----- 136  
Db 2044 ASDPSRKVAKPRLKPIWDREAMKTTIKVGNDEVPDVPVGEPPPKKEWIFNEKFPVDD 2103  
QY 137 -----GA-H-----TI-A-N----- 143  
Db 2104 QKRIESEDYKTRFVLRGATRKAGLYTLTATNAGSDKHSVEVIVLGKPSPLGLEVS 2163  
QY 144 N-----H-----TDM-----ME----- 150  
Db 2164 NVYEDRADLEWKVPEDDGGAPIDHYEIEKMDLATGRWVPCGRSETTKTTPNLPQGEYK 2223  
QY 151 V-----D-----G-----D-----V 155  
Db 2224 FRVAVNKEGESDPLTTNTAILAKNPYEVPGKVDKPELVWDKDHVDLAWNAPDDGGAPI 2283  
QY 156 E--I-----P-----P-----NKA-----V 163  
Db 2284 EAFVIEKKDKNRWEALVVGQKQATVPNLKEGEYQFRISARNKAGTGDSPSDRV 2343  
QY 164 V-----L-----R-----G-----H-----ES----- 170  
Db 2344 VAKPRNLAPRIHREDLSDTTVKVGATLKFIHVHIDGEPAPDVTWSFNGKGIGESKAQIENE 2403  
QY 171 -----F-----I-----C-----EV----- 172  
Db 2404 PYISFALPKALRKQSGKYTTITATNINGDTSVINIKVSKTPKPKPIEVDVDFEDRAT 2463  
QY 173 -----F-----I-----C----- 175  
Db 2464 LDWKPPDDGGEPIEFIEIEKMTKGIWVPCGRSGDTHFTVDSLNGKHDIYKFRVAVNS 2523  
QY 176 -----G-----Q-----HK-G-----P-----I----- 269

Db 2524 EGSPDPLETETDILAKNPPDRPDRPGRPEPTDWDSDHVDLKWDPPLSDGGAPEEYQIEK 2583  
QY 177 -----W----- 177  
Db 2584 RTKYGRWEPAITVPGQTTATVPLDTPNEBYEVRVAVNKGSPSPDSDASKAVIAKPRNL 2643  
QY 178 -----N-----PVS-----D-----L----- 183  
Db 2644 KPHIDRALKNLTIKAGQISFPDVPVGBEPAPTWTWHPDNREIRNGRGRVCLDNPEYQSK 2703  
QY 184 L-----A-----S-G-----S-GD----- 190  
Db 2704 LVVKQMERGDSGTFTIKAVNNGEDEATVKINVIDKPTSPNGPLDVSVDHGDHVTLLNWA 2763  
QY 191 -----S-----TA-----R----- 194  
Db 2764 PDDGGIPIENYVIEKYDTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVAVNAEAGE 2823  
QY 195 -----I----- 195  
Db 2824 SDPLETFTTAKDPDFPKTNAPETDWDKDHVDLEWKPPANDGGAPIEYVEMKDE 2883  
QY 196 ---W-----NL-----S-E----- 200  
Db 2884 FSPFNDVAHVPAQOTNATVGNLKEGKYEFRIRAKNKGAGLGDPSDSASAVAKARNVPPV 2943  
QY 201 ---NS-----T-----S-----GST----- 207  
Db 2944 IDRNSTQIEIKVQAGQFSLNIPVSGEPTPTITWTFEGTPVESDDRMKLNNEDGKTFHVK 3003  
QY 208 Q-L-----VL-----R----- 212  
Db 3004 RALRSTGTVIIKAENENGTDIAEVKTVLDHPSSPRGPLDVTNIVKDCDLAWKSPEDD 3063  
QY 213 -----H-----C----- 214  
Db 3064 GGAEISHYVIEKDAATGRWTACGESKDTNFHVDLDTQGEYKFRVAVNRHGDSDPLEA 3123  
QY 215 ---I-----R-----EG----- 218  
Db 3124 REALIAKDPFDRADKPGTPEIVDMKDHADLKWTTPPADDDGAPIEGYLVEMRTPSGDWVP 3183  
QY 219 ----- 218  
Db 3184 AVTVGAGELTATVDGLKPGQTYQFRVAVNKAAGESTPSPSRMTWAKPRHLAPKINRDMF 3243  
QY 219 ---GQ-----DV-----P----- 223  
Db 3244 VAQRVKAGQTLNFDVNVGEPAPKIEWFLNGSPLSSGGNTHIDNNTDNNKLTTKSTARA 3303  
QY 224 -----S-----N-----KQVT-----S-----LD----- 232  
Db 3304 DSGKYKIVATNESGKDEHVDVNIIDIPGAPEGLRHDKITKESVVLKWDPELDDGSGPI 3363  
QY 233 -----W-----N-SE-----GT-----L----- 239  
Db 3364 TNYVVEKQSDGGRWVPCGETSOTSLKVNKLSGHEYKFRVAVNRQGTSAPLTSDHAIYA 3423  
QY 240 ---L-----A-----T-----G-----SY-----DG 247  
Db 3424 KNPFDEPDAPTDTVTPVDWKDHVDLEWKPPANDGGAPIDAYIVEKDKFGDWVECARVDG 3483  
QY 248 -----F----- 251  
Db 3484 KTTKATADNLTPGETTYQFRVAVNKAAGPKSPDPTGNVVVAKPRMAPKLNLAGLLDRIK 3543  
QY 252 -----W-----TKD-G-----NL-----A-----S-----T-L----- 262  
Db 3544 AGTPIKLDIAPGEPAPVAKWKANDATIDTGARADVNTPTSSAHIHFAVRGDTGVYKI 3603  
QY 263 -----G-----Q-----HK-G-----P-----I----- 269

Db 3604 I V E N E H G K D T A Q C N V T V D V P G T P E G L K I D E I H K E G C T L N M K P P T D N G G T D V L H I V E K 3663  
QY 270 - - - - - F - - - - - A - - - - - L - - - - - 272  
Db 3664 M D T S R G T W Q E V G T F P D C T A K N K L V P G K E Y A F R V A V N I Q G S K P L E A E S P I I A K N Q P D V 3723  
QY 273 - - - - - K W N - - - - - K K - - - - - G - - - - - 278  
Db 3724 P D P V D K P E V T D M D K R I D I K W N P T A N N G A P V T G Y I V E K K E G S A I W T E A G K T P G T F S A 3783  
QY 279 - N - - - - - F - - - - - T L - - - - - S - - - - - A G - - - - - 285  
Db 3784 D N L K P G V E Y F R V I A V N A A G S P D S P D T P Q I T K A R Y L K P K I L T A S R K I K I K A G F T H N L E 3843  
QY 286 - - - - - V D - K - T T - - - - - 290  
Db 3844 V D F I G A P D P T A T W T V G D S G A A L A P E L L V D A K S T T S I F P F S A K R A D S G N Y K L K V N E L G E 3903  
QY 291 - - - - - I - - - - - W - - - - - D - - - - - 294  
Db 3904 D E A I P E V I Q D R P S A P E G P L E V S D V T K D S C V L N M K P P K D G G A E I S N Y V V E K R D T K T N T W 3963  
QY 295 - - - - - A - - - - - H - - - - - T - G - - - - - E - - - - - A K O Q F - - - - - 304  
Db 3964 V P S A F V T G T S I T V P K L T E G H E Y F R V M A E N T F G R S D S L N T D E P V I A K D P F G T P G K P G R P 4023  
QY 305 - - - - - A - - - - - Q - - - - - S - - - - - 306  
Db 4024 E I V D T D N D H I D I K W D P P R D N G S P V D H Y D I E R K D A K T G R W I K N T S F V O G T A F S D T R V O K 4083  
QY 307 - H - - - - - S A - - - - - P - - - - - A - - - - - L D - - - - - V - - - - - 314  
Db 4084 G H T Y E Y R V V A N K A G P Q S D S A A T A K P M H E A P K F D L D L D G K F R V K A G E L V I T I P F 4143  
QY 315 - - - - - D - - - - - W - - - - - Q - - - - - S - - - - - 318  
Db 4144 T A S P Q D I S W T K E G K P L A G V E T T D S Q T K L V I P S T R R S D S G P V K I K A V N P Y G E A E A N I K I 4203  
QY 319 - - - - - N - - - - - N - - - - - 320  
Db 4204 T V I D K G A P E N I T Y P A V S R H C T L N W D A P K D G G A I A G Y K I E Y Q E V G S Q I W D K V P L G I S 4263  
QY 321 - - - - - T - - - - - F - - - - - A - - - - - S - - - - - C - - - - - S T - - - - - 327  
Db 4264 G T A Y T V R G L E H G Q O Y F R I A E N A V G L S D Y C Q G V P V I K D P P D P P C A P S T P E I T G Y D T N Q 4323  
QY 328 - - - - - D - - - - - M - - - - - C I - - - - - 331  
Db 4324 V S L A M P P R D D G S P I L G Y V V E R F E K R G G D W A P V M P M V K G T E C I V P G L H E N E T Y Q P R V 4383  
QY 332 - - - - - H - - - - - V - - - - - C - - - - - 334  
Db 4384 R A V N A A G H E P S N G S E P V T C R P Y V E K P G A D A P R V G K I T K N S A E L T W N R P L R D G G A P I D G 4443  
QY 335 - - - - - K L G - - - - - 337  
Db 4444 Y I V E K K L G D N W T R C N D K P V R D T A F E V K N L G E K E E Y F R V I A V N S A G E G E K S P S D I V L 4503  
QY 338 - - - - - Q - - - - - D - - - - - R - - - - - P I - - - - - 342  
Db 4504 I B E Q P G R P I F D I N N L K D I T V R A G E T I Q I R I P Y A G N P K P I I D L F N G S P I F E N E R T V D V 4563  
QY 343 - - - - - K - - - - - T - - - - - F - - - - - Q - - - - - 346  
Db 4564 N P G E I V T T G S K R S D A G P Y K I S A T N Y K G O T C K L A N F V L D A P G K T G P I R A T D I Q A D A M 4623  
QY 347 - - - - - G H - - - - - T - - - - - N - - - - - E - - - - - V - - - - - N - - - - - 353  
Db 4624 T L S W R P P K O N G D A I T N Y V V E K T P G D W V T G H P V G T I L R V N L D A N T P Y F R V R A E N Q 4683  
QY 354 - - - - - A I - - - - - K - - - - - W - - - - - D - - - - - P - - - - - 359  
Db 4684 Y G V G E P L E T D D A I V A K N P F D T P G A P Q E A V E T S E B A I T L Q W T R P T S D G A P I Q G Y V I E K 4743

QY 360 - - - - - T - - - - - G N - - - - - 362  
Db 4744 R E V G S T E W T K A A F G N I L D T K H R V T G L T P K K T Y E F R V A A Y N A A Q G E Y S V N S V P I T A D N A P 4803  
QY 363 - - - - - L - - - - - 363  
Db 4804 T R P K I N M G M L T R D I L A Y A G E R A K I L V P F A A S P A P K V T F S K G E N K I S P T D P R V K V E Y S D F L 4863  
QY 364 - - - - - L - - - - - A S - - - - - C - - - - - 367  
Db 4864 A T L T I E K S E L T D G G L Y F V E L N S Q S D S A S I R L K V V D K P A S P O H I R V E D I A P D C C T L Y W M 4923  
QY 368 - - - - - S D - - - - - D M - - - - - T - - - - - 372  
Db 4924 P F S S D G G S P I T N Y I V E K L D L R H S D G K W E K V S F V R N L N Y T V G L I K D N R Y R F R V R A E T Q Y 4983  
QY 373 - - - - - L - - - - - K - - - - - I - - - - - W - - - - - S M K - - - - - Q - - - - - 380  
Db 4984 G V S E P C E L A D V V A K Y Q F E V P N Q P E A P T V R D K D S T W A E L E W D P P R D G S - K I I G Y Q V Q Y R 5042  
QY 381 D - - - - - N - - - - - C V H D - - - - - L - - - - - Q - - - - - 389  
Db 5043 D T S S G R W I N A K M D L S E Q C - H A R V T G L R Q N G E F E F R I I A K N A A G F S K P S P P S E R C Q L K S R F 5101  
QY 390 - - - - - H - - - - - N - - - - - K - - - - - E I - - - - - Y - - - - - T I K W - - - - - S - - - - - 400  
Db 5102 G P P G P I H V G A K S I G R N H C T I T W A P L E D G G S K I T G Y N V E I R E Y G S T L - W T V A S D Y N V R E 5160  
QY 401 P - - - - - T - - - - - G P - - - - - G - - - - - T - - - - - N - - - - - 407  
Db 5161 P E F T V D K L R E F N D Y E F R V V A I N A A G K I P S L P S G P I K I Q E S G G S R P Q I V V K P E D T A Q P Y N 5220  
QY 408 - - - - - N - - - - - P - - - - - 409  
Db 5221 R R A V T C E A V G P E P T A R W L R N G R E L P E S S R Y R F E A S D G Y K F T I K E V W I D I D A G E Y T V E V 5280  
QY 410 - - - - - A N L - - - - - 413  
Db 5281 S N P Y G S D T A T A N L V Q A P P V I E K D V P N T I L P S G D L V R L K I Y P S G T A P F R H S L V L N R E E I D 5340  
QY 414 M - - - - - L - - - - - A - - - - - S - - - - - F - - - - - 420  
Db 5341 M D H P T I R I V E F D D H I L I T I P A L S V R E A G Y E V T V S N D S G E A T F G M L N V T G L P E A P Q G P L 5400  
QY 421 - - - - - D - - - - - S - - - - - T V - - - - - 424  
Db 5401 H I S N I G P S T A L S W R P P V T D G S K I T S Y V E K R D L S K D E W V T V T S N V K D M N Y I V T G L F E N 5460  
QY 425 - - - - - 424  
Db 5461 H E Y E F R V S A Q N E N I G A P L V S E H P I I A R L P D P P T S P L N L E I V O V G D Y V T L S W O R P L S D 5520  
QY 425 - - - - - R L - - - - - W - - - - - 427  
Db 5521 G G R L R G Y I V E K Q E B E H D E W F R C N Q P S P P N V N V P N L I D G R K Y R Y R F A V N D A G L S D L A 5580  
QY 428 - - - - - D - - - - - V - - - - - D - - - - - R - - - - - 431  
Db 5581 E L D Q T L F Q A S G S G E G P K I V S P L S D L N E E V G R C V T F E C I S G S P R P Y R W F K G C K E L V D T S 5640  
QY 432 - - - - - G - - - - - I - - - - - C - - - - - H - - - - - 436  
Db 5641 K Y T L I N K G D K V L I I N D L T S D D A D E Y T C R A T N S S G T R S T R A N L R I K T K R V F I P P K V H G G 5700  
QY 437 - - - - - T L - - - - - 438  
Db 5701 Y E A Q G E T I E L K I P Y K A Y P Q G E A R W T K G E K I E N N S K F S I T T D D K F A T L R I S N A S R E D Y G 5760  
QY 439 - - - - - 438  
Db 5761 E Y R V V V E N S V G S D S G T V N V T V A D V P P P R P P I I E N I L D E A V I L S W K P P A L D G G S L V T N Y T 5820



QY 12 --VYR-----Y-----L-----16  
Db 912 QEIYRTQAVLTCKVNTPRAPLVWYRGSKAIOEGDPRFIEIKDAVGRCTLTIKEVEDDQ 971  
QY 17 -----Q-----ES-----19  
Db 972 AEWTAIRITQDVFSKVQVYVEPRHTFVPMKSKQVNESDLATLETVDNDKDAEVMWHDG 1031  
QY 20 --G--F--S-----H-----S-----A--F--26  
Db 1032 KRIDIDGVFKVBSNRKRRLIINGARIEDHGEYKCTTKDDRTMAQLIIVDAKKNFIVALK 1091  
QY 27 -----T--F--G--I-----K--S-----H-----33  
Db 1092 DTEVIEKDDVTLMQTKDTKPGIWRNGKQISSMPGKFETQSRNGTHTLKIGKIEMNE 1151  
QY 34 ----ISO-----S-----N-----38  
Db 1152 ADVYEDQAGLRGSCNVTLEAEKRPILNWKPKKIEAKAGEPCVVKVPFIKGTTRGDPK 1211  
QY 39 --I--NG-----N-----AL-----43  
Db 1212 AQLKNGKPIDEMRKLVEVIIKDDVABIVFKNPQLADTGKWALELGNAGTALAPFELF 1271  
QY 44 V-----PP-----AA-----LI--S-----II--Q-----KG 56  
Db 1272 VKDKPKPKPLTKNVTABGLDLVNGTDPDPDEGAPKAVYIIEMQGRSGNMAKVGETKG 1331  
QY 57 -----L--QY--VEA-----B--V-----SI--N-----67  
Db 1332 TDFVKVDLKEHGEYKFRVKALNEGLSLDTGESVLAKNPYGVPGPKMNDALDVKDHC 1391  
QY 68 -----EDG-----T-----L-----72  
Db 1392 TLAWEPEDGGAPITGYIIERREKSEKWHQVQGTQKPCCELTDKKVVEDKEYLYRVKA 1451  
QY 73 -----F-----D-----74  
Db 1452 VNKAGPGDPCDHGKPIKMAKAKASPEFTGGIKDLRLKVGETIKYDVPISGEPLPECLWV 1511  
QY 75 --GRP-----IE-----S-----L-----S-----82  
Db 1512 VNGKPLKAVGRVKMSSERGHMKIENAVRADSGKFTITLKNSSGSCDSTATVTVVGRPT 1571  
QY 83 -----L--I-----D--AVM-----PD-----VV--Q-----T--94  
Db 1572 PPKGPLDIADVADGATLSWNPPDDGDPPLTGYIIVEAQMDNKGKIIIEVGKVDNPTTL 1631  
QY 95 -----R-----Q--Q-----Q-----97  
Db 1632 KVNGLRNGNYKFRVKAVNNEGESEPLSADQYTIKDPWDEPGKGRPEITDFDADRIDI 1691  
QY 98 -----A--Y-----RD-----101  
Db 1692 AWEPPHKGGAPEEYIEVVRDPTKEMKEVVRVPTDNASISGLKEGEYQFRVRAVKA 1751  
QY 102 -----K--LA-----Q-----105  
Db 1752 GPGQSPSEKQLAKPKFIPAWLKHDNLKSIITVKAGATVRWEVKIGGEPIEYKFKGNQ 1811  
QY 106 Q-----Q-----A-----A-----A-----110  
Db 1812 QLENGIQLTIDTRKNEHTILCIPSAMRSDVGEYRLTVKSHGADESKANLTVLDRPSKN 1871  
QY 111 -----A-----A-----A-----A-----113  
Db 1872 GPLEVDFEDNLNLSWKPPDDGGEPIEYVEVEKLDATGTRWVPCAKVKDKAHDGLK 1931  
QY 114 -----A-----A-----A-----A-----118  
Db 1932 KGQTYQFRKAVNKEGASDALSTDKTKKANPNYDEPKGTGTDPDWDADRVSLWEPPK 1991  
QY 119 S-----Q-----Q--G--S-----A-----K--N-----GE 128

Db 1992 SDGGAPITQYVIEKKGKGRDWEQCKVSGDQTNABILGLKEGEYQFRVKAIVKAGPGE 2051  
QY 129 -----N-----T--AN-----GE-----E--N-----136  
Db 2052 ASDPSKRVKAPRNLPKWDIREAMKTTITIKVGNDFEFDVVRGEPKKEWIFNEKPVDD 2111  
QY 137 -----GA--H-----TI--A--N-----TDM-----ME--150  
Db 2112 QXIRIESEDYKTRFVLRGATRKHAGLYTLTATNASGDKHSVEVILGKPSPLGLEVS 2171  
QY 144 N-----H-----TDM-----150  
Db 2172 NVYEDRADLEWKVPEDDGGAPIDHYIEKMDLATGRWVPCGRSETTKTTVPNLQPGHEYK 2231  
QY 151 -----V--D-----G-----D-----155  
Db 2232 FRVRANKGESDPLTTNTAILAKNPYEPGVKVDKPELVDWKDHDVLDLAWNAPDDGAPI 2291  
QY 156 E--I-----P-----P-----NKA-----V 163  
Db 2292 EAFVIEKKDKNGRWEALVVPDQKTATVPNLKEGEYQFRISARNKAGTGDPSPSDRV 2351  
QY 164 V-----L--R-----G-----H-----ES-----170  
Db 2352 VAKPRNLAPRIHREDLSDTTVKVGATLKFIHVHIDGEPAPDVTWFSNKGKIGESKAQIENE 2411  
QY 171 -----172  
Db 2412 PYISRPALPKARKOSGKYTTATNINGTDSVTINIKVSKPKTKPGPIEVTDFEDRAT 2471  
QY 173 -----F-----I--C-----175  
Db 2472 LQWKPPDDGGEPIEYIEKMKTKDGIWVPCGRSGDTHFTVDSLNGDHYKFRVKAIVS 2531  
QY 176 -----176  
Db 2532 EGSPDPLETETOILAKNPDRPDRPGRPEPTOWDSHDVLDKMDPPLSDGGAPIEYQIEK 2591  
QY 177 -----W-----177  
Db 2592 RTKYGWEPATITVPGGQTTATVPDLTPNBEYFRVVAVNVKGGSPSDSKAVIAKPRNL 2651  
QY 178 -----N-----PVS-----D-----L-----183  
Db 2652 KPHIDEDALKNLTIKAGQISFDVPVSGEPAPTIVHWPDRNREIRNGRVLKDNPEYQSK 2711  
QY 184 L-----A-----S--G-----S--GD-----190  
Db 2712 LVVKQMERGDSGTFTIKAVNANGEDATVKINVIDKPTSPNGPLDVSVDVHGDHVTLNWRA 2771  
QY 191 -----S-----TA-----R-----194  
Db 2772 PDDGGIPIENYVIEKYDTASGRWVPAKAGDKTTAVVDGLIPGHEYKFRVAANAGE 2831  
QY 195 -----195  
Db 2832 SDPLETFGTTLAKDPDKGKTNAPEITDWDKHDVLEWKPANDGAPIEYVVEWKDE 2891  
QY 196 -----W-----NL--S--E-----200  
Db 2892 FSPFMDVAHVPAQTNAFTVGNLKEGSKYEFIRAKNKAAGLGDPSDSASAKARNVPPV 2951  
QY 201 ---NS-----T--S-----S-----GST-----207  
Db 2952 IDRNSIQIBIKVKAQDPSLNI PVSGEPTPTITWTPEGTVESDDRMKLANNEDGKTFHVK 3011  
QY 208 Q--L-----VL-----R-----212  
Db 3012 RALRSDTGYIIKAENENGTDAEVKVTVDHPSSPRGFLDVTNIVKDCGLAWKEPDD 3071  
QY 213 -----H-----C-----214

Db 3072 GBAISHYVIEKODAAATGRWTACGESKDTNFHVDLTCQHEHYKFRVAVNRHGDSDPLEA 3131  
QY 215 ---I---R---EG--- 218  
Db 3132 REALIAKDPDRADKPGTPEIVDMKDHADLKWTTPADDDGAGPIEGYLVEMRTPSGDWVP 3191  
QY 219 --- 218  
Db 3192 AVTVGAGELTATVDGLKPGQYQFRVKALNKAGESTPSDSRMTWAKPRHAPKINRDMF 3251  
QY 219 ---GO---DV---P--- 223  
Db 3252 VAORVAGQTLNFVNVEGEPAPKIEWFLNGSPSSGCGNTHIDNNTDNNTKLTAKSTARA 3311  
QY 224 ---S---N---KQVT--S---LD--- 232  
Db 3312 DSGYKIVATNESKDBHEVDVNIIDIPGAPGELRHKIDTKGSVLWKWDEPLDDGGSPI 3371  
QY 233 ---W---N---SE---GT---L--- 239  
Db 3372 TNYVVEKQEDGGRWVPCGETSDTSLKYNKLSGHEHYKFRVAVNRQGTSAPLSDHAIVA 3431  
QY 240 ---L---A---T---G---SY---DG 247  
Db 3432 KNPFDEPDAPTPTVDWDKDHVLEWKPANDGGAPIDAYIIVEKKDKFGDWVECARVDG 3491  
QY 248 ---F--- 251  
Db 3492 KTKATADNLTPGETYQFRVAVNAKAGPGKPSDPTGNVAVKPRMAPKLNLAGLLDLRIK 3551  
QY 252 ---W---TKD-G---NL---A---S---T-L- 262  
Db 3552 AGTPIKLDIAFEGEPAPVAKWKANDAIDTIDGARADVNTTSSAIHIFSAVRGDTGYKI 3611  
QY 263 ---G---Q---HK-G---P---I--- 269  
Db 3612 IVENEHGKTAQCNTVTVLDVPGTPEGPKIDIEIHKEGCTLNWKPPTDNGGTDVLHVIVEK 3671  
QY 270 ---F---A---L--- 272  
Db 3672 MDSRGTWQEVGTPDCTAKNKLVPGEYAFRVAFAVNAVLOGESKPLEAEPIIAKQFDV 3731  
QY 273 ---KWN---KK---G--- 278  
Db 3732 PDVVKPEVTDWKORIDIKNWPTANNGAPVTGYIIVEKKEKSAIWEAGKPTGTTFA 3791  
QY 279 -N---F---IL--S---AG--- 285  
Db 3792 DNLKPGVEYEFRIAUNAAGSDPSDTPQITKARYLKPILKTASRKIKIKAGFTHNLE 3851  
QY 286 ---VD-K--TT--- 290  
Db 3852 VDFIGAPDPTATWVDSGAALAPELLVDKASSTTSIFPPSAKRAUSGNYKLVKNELGE 3911  
QY 291 ---I---I---W---D--- 294  
Db 3912 DEALFEIVQDRSAPGPLEVSDVTKSCVLNWKPKPDGGAEISNYVVEKEDTKNTW 3971  
QY 295 ---A---H---T-G---E---AKQOF--- 304  
Db 3972 VPVSFAVTGTSITVPKLTEGHEYEFRVMAENTFGSDSLNTBFLAKDPFGTPGKGRP 4031  
QY 305 ---P---F--- 306  
Db 4032 EIVDTDNHDHIDIKWPPRNGGSPVDHYDIERKDAKTGRWIKNTSPVQGTAFSDTRVQK 4091  
QY 307 -H---SA---P---A---LD---V--- 314  
Db 4092 GHTYEYRVAVNAKAGCPQSDSSAAATAKPMHEAPKFDLDDGKEFRVKAPELVITIPP 4151  
QY 315 ---D--W---Q---S--- 318  
Db 4152 TASPQPDISWTKEGKPLAGVETTDSTQTLVIFSTRSDSGPVKIKAVNPYGEAEANIKI 4211

QY 319 ---N---N--- 320  
Db 4212 TVIDKFGAPENITYPAVSRTCTLNWDAPKDDGAGIAGYKIEYQVGSQIWDKVPGLIS 4271  
QY 321 ---T---F---A---S-C---ST--- 327  
Db 4272 GTAYTVRGLEHGQYFRIRAEANAVGLSDYCOGQVVPVVIKPPDPGAPSPTEITGYDTNQ 4331  
QY 328 ---D---M---CI--- 331  
Db 4332 VSLAMNPRDDGGSPILGYVVERFEKRGGDWAPVMPVMVKGTECIVPLHENETYQFRV 4391  
QY 332 ---H---V-C--- 334  
Db 4392 RAVNAAGHSPSGSEPVTCRPYVEKPGAPDAPRVGKITKNSAELTWRNRLRGGAPIDG 4451  
QY 335 ---KLK--- 337  
Db 4452 YIVEKKLGDNDWTRCNDKPVRTAFEVKNLGESEYEFRVIAVNSAGEGEPKPSDLVL 4511  
QY 338 ---Q---D---R---PI--- 342  
Db 4512 IEEQGRPIFDINNLKDIIVRAGETIQIRIPYAGGNPKFIIDLFGNSPIFENERTVVDV 4571  
QY 343 ---K-T---F---O--- 346  
Db 4572 NPGEIVITTTGSKSDAGPYKISATNKGDKTCKLVFVLDAPKGTGPIRATDIQADAM 4631  
QY 347 ---GH---T---N---E-V--N- 353  
Db 4632 TLSWRPPKONGDAITNYVVEKRTPGGDWTVGHVPGTTLRVNLDANTPYEFRVAENQ 4691  
QY 354 ---AI--K---W---D--P- 359  
Db 4692 YGVGEPLETDDAIVAKNPFDTPGAPQPEAVETSEAITLQWTRPTSDGAPIQGYVIEK 4751  
QY 360 ---T---GN--- 362  
Db 4752 REVGSTEWTKAAGFNILDTKHRVTGLTPKTYEFRVAAVNAAGQGEYSVNSVPIADNAP 4811  
QY 363 ---L--- 363  
Db 4812 TRPKINMGMLTRDILAYAGERAKILVFAASPAPKVTFSKGENKISPTDPRVKVEYSDFL 4871  
QY 364 ---L---AS---C--- 367  
Db 4872 ATLTIKSELTDGGLYFVLENSQGSDSASIRLKVVDKASPQHIRVEDIAPCCTLYWM 4931  
QY 368 ---SD---DM---T--- 372  
Db 4932 PFSSDGGSPITNYIVEKLDLRHSDGKWEKVSFVRNINVTVGGLIKDNRYRFRVRAETQY 4991  
QY 373 ---L---K---I---W---SMK---Q-- 380  
Db 4992 GVSEPCELADVVKVAKYQFVNPQPEAPTVRDKDSTWAELEWDPRDGS-KIIGYQVQYR 5050  
QY 381 D-----N-----CVHD---L-Q-----Q----- 389  
Db 5051 DTSSGRWINAKMDLSEQC-HARVTGLRQNGEPEFRIIAKNAAGFSKPSPPSECOLKSFR 5109  
QY 390 ---H---N---K---EI--Y--TIKW--S--- 400  
Db 5110 GPPGPIHVGAKSIGNHCTITWMAPLEDGGSKITGYNVEIREYGSTL-WTVASDYNVRE 5168  
QY 401 P--T-----GP-----G-----T---N 407  
Db 5169 PEFTVDKLEFNDYFRVAVNAAGKIPSLPSGPIKIQESGGRPQIVVVKPDTAQPN 5228  
QY 408 ---N---P----- 409  
Db 5229 RRAVFTCEAVGRPEPTARWLNRGRELPESSRYRFEASDGYKFTIKEVWDIDAGEYTV 5288

QY 410 -N-----ANL----- 413  
Db 5289 SNFYSGDTATANLVQAPPVIEKQDVNTILPSGLVRLKIYFSGTAPFRHSLVNLNEED 5348  
QY 414 M-----L-----A-S-A-----S-----F----- 420  
Db 5349 MDHPTIRIVEFDHILITIPALSVREAGRYEYTVSNDSGEATTFGLNVTGLPEAPQGPL 5408  
QY 421 -----D-S-----TV----- 424  
Db 5409 HISNIGSTATLSWRPVTGGSKITSYVVEKEDLSKDEWVTYSNVKDMYIIVTGLPEN 5468  
QY 425 ----- 424  
Db 5469 HEYEFVRSQNGENIGAPLVSEHPHIIARLPDPPTPLNLEIVQVGVDYVTLWSQRPLSD 5528  
QY 425 -RL-----W----- 427  
Db 5529 GGGRLRGYIVKQEBEHEWFRNCQNPSPNNVNPVNLIDGRKYRYRFAVNDAGLSDLA 5588  
QY 428 --D-----V-----D-----R----- 431  
Db 5589 ELDQTLFQASGSGEGKIYSPSLDLNEEVGRCVTFECSISGSPRPRYRWFKGCKELVDTS 5648  
QY 432 -----G-----I-----C-----I-----H----- 436  
Db 5649 KYTLINKGDKQVLIINDLTSDDADEVTCRATNSSGTRSTRANLRKIKTKPRVFIPPKYHGG 5708  
QY 437 ----- 438  
Db 5709 YEAQGETIELKIPYKAYPQGEARWTKGGEKIENNSKFSITDDKPFATLRISNASREDYG 5768  
QY 439 ----- 438  
Db 5769 EYRVVSVSGSDGTNTVTADVPRPPPIIENILDEAVILSWKPPALDGGSLVNTYT 5828  
QY 439 -----T-----KH-Q-----E-----PV----- 445  
Db 5829 IEKREAMGSGWSPCAKSRVYTTIEGLRAGQVEFPRIIAENKHGQSKPCEPTAPVLIPGD 5888  
QY 446 -----Y-----S-----V----- 448  
Db 5889 ERKRRRGYDVEGKIVRGKGTVSSNYDNYVDFIMQYYPQPVVEIKHVDLHYDIHEEL 5948  
QY 449 -AF-----S-----P----- 452  
Db 5949 GTCAFGVHRVTERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAPED 6008  
QY 453 D-----G-----R-----Y----- 456  
Db 6009 DNEWMVIEBFMSGGELFEKVADEHNKWSDEAVEYMRQVCKGLCHMHENNYVHLDLKPEN 6068  
QY 457 -----L-----A----- 458  
Db 6069 IMFTTKRSNELKLIDFGLTAHLDPKOSVKVTTGTAEFAAPEVAEGKPVGYTDMWSVGL 6128  
QY 459 -SG-S-F-----D-----K-C----- 466  
Db 6129 SVILLSGLSPFGGENDDETLRVKSCDWMDDSAFSGISGDKDFIRKULLADPNTRMTY 6188  
QY 467 H-----I-----W-----N----- 470  
Db 6189 HQALEHPMLTPGNAPGRDQIPSSRYTKIRDSIKTKYDAWPEPLPPLGRISNVSSLRKH 6248  
QY 471 -----TQ-----T-----G-A-L-----V-H----- 478  
Db 6249 PQEYSIRDADFDRSEAPRFIVKPYGTEVGEOSANFYCRVIASSPPVVTWHKDDRELKQ 6308  
QY 479 S-Y-R-----G-T-----G-----G-----I-----PE----- 488  
Db 6309 SVKMKRYNGNDYGLTINRVKGDGGEYTVRAKNYSYGTKEEIVFLNVTNRHSBPLKPEPLE 6368  
QY 489 -----V-----C-----W----- 491

Db 6369 PMKADSPRPVBEFKEKERRGAPFFHLNRNLIQNHQCKLCTSLQGNPNPTIEWMKDGH 6428  
QY 492 -----NA-----AG-----D-----K-----V 498  
Db 6429 VDEDRVQVFRSGVCSLEIFNARVDDAGTYVTATNDLGVVDVSECVLTVQTKGGEPIPRV 6488  
QY 499 -----G-----A-----SA----- 502  
Db 6489 SSFRPRAYDTLSTGTDVDRSHSYADMRRSLRDVSPDVSAAADDLKTITNELPSFTA 6548  
QY 503 --SD-----GS-----V----- 507  
Db 6549 QLSDETSEVGSAAEFAAAGVSGOPELIEWLHNGERISESDSRFRASVYVAGKATLRISDAK 6608  
QY 508 -----C-----VL-----D-----LR--K 514  
Db 6609 KSDEGYLCRASAGQEQTATLTVKGDQPLLNHAGQAVESELRTVK 6657

RESULT 12  
T27935  
hypothetical protein ZK617.1b - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T27935; T28031  
R:White, S.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: Z20442  
A:Accession: T27935  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-7160 <W1>  
A:Cross-references: UNIPROT:Q23551; EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617.1b  
A:Experimental source: clone ZK617  
R:Harris, B.  
submitted to the EMBL Data Library, May 1996  
A:Reference number: Z20458  
A:Accession: T28031  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-7160 <W1>  
A:Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.1b  
A:Experimental source: clone ZK829  
C:Genetics:  
A:Gene: CESP:ZK617.1b  
A:Map position: 4  
A:Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 597/3; 3067/1; 3141/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3  
C:Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology; I

Query Match 70.3%; Score 2572.2; DB 2; Length 7160;  
Best Local Similarity 8.0%; Pred. No. 5.6e-52;  
Matches 472; Conservative 29; Mismatches 10; Indels 5358; Gaps 394;

QY 1 M-----S-----I-----S-S-----D----- 6  
Db 1113 MTEAERRQLFPGKKVEKWDIPLPEKTVQQQVDKICEWKCTYSRPNKIRWYKDKRKEIFS 1172  
QY 7 -----EV-----N-----PL----- 11  
Db 1173 GGLKYKIVIEKNVCLTIINNPEVDVTGKYTCBANGVPTHAQLTVLEPPMKYSFLNPLNT 1232  
QY 12 --VYR-----Y-----L----- 16  
Db 1233 QBIYRTKQAVLTCKVNTPRAPLVWYRGSKAIQEGDPRFIEKDAVGRCTLTKEVEDDQ 1292  
QY 17 -----Q-----ES----- 19  
Db 1293 AEWARTITQDVFSKVQVVEPRHPTFVPMKSCQKNESDLATLETVDNDKDAEIVVWHDG 1352  
QY 20 -----G-F-----S-----H-----A-----F----- 26

Db 1353 KRIDIDGVKPKVSSNRKRRLIINGARIEDHGEYKCTTKDORTMAQLIVDAKNKFI VALK 1412  
QY 27 -----T-----F-G-I-----K-----S-----H----- 33  
Db 1413 DTEVIEKDDVTLMCQTKDTKPGIWFNRNGKOISSMPCGKFETQSRNGHTLTKIGKIEWNE 1472  
QY 34 -----ISQ-----S-----N----- 38  
Db 1473 ADVVEIDQAGLRGSCNVTVLEAEKRPILNWKPKKIEAKAGEPCVVKVFFQIKGTRRGDPK 1532  
QY 39 -----I-----NG-----AL----- 43  
Db 1533 AQILKNGKPIDDEEMRKLVEIHKDDVAEIVFNKPNQADTKGWALELNGSAGTALAPPELF 1592  
QY 44 V-----PP-----AA-----LI-----S-----II-----Q-----KG 56  
Db 1593 VKDKPKPKPLETKNVTAEGLDLVWGTPDPDEGAPVKAVIIEMOGRSGNWAUVGETKG 1652  
QY 57 -----L-----QY-----VEA-----E-V-----SI-----N----- 67  
Db 1653 TDFKVKDLKEHGEYKFRVKALNECGLSDPLTGESVLAKNPGVGPCKPKNMDAIDVDKXHC 1712  
QY 68 -----EDG-----T-----L----- 72  
Db 1713 TLAWEPPEEDGAPITGVIIERREKSEKWHQVGQTKPDCCELTDKKVVEDKEYLYRVKA 1772  
QY 73 -----F-----D----- 74  
Db 1773 VNKAGDPDCHGPKPKAKKAKASPEFTGGIKDLRLKVGETIKYDVPISGEPLPECLMW 1832  
QY 75 -----GRP-----IE-----S-----L-----S----- 82  
Db 1833 VNGKPLKAVGRVKMSSEKHKIMKIENAVRADSGKFTITLKNSSGSCDSTATVTVGRPT 1892  
QY 83 -----L-I-----D-AWM-----PD-----VW-Q-----T- 94  
Db 1893 PPKGPLDIADVCDATLSWNPDDGDDPLTGVIEAQMNDKNGKIEVGVKVDPNNTTL 1952  
QY 95 -----R-----Q-Q----- 97  
Db 1953 KVNGLRNKGNKFRVKAVNNEGESEPLSADQYQIKDPWDEPGKPRPEITDFADRIDI 2012  
QY 98 -----A-----Y-----RD----- 101  
Db 2013 AWEPPHKDGAPEIEYIEVEVRDPTKEWKEKRVDPDTNASI SGLKEGKEYQFRVAVNKA 2072  
QY 102 -----K-LA----- 105  
Db 2073 GPGQPSPEBKOLAKPKFIPAWLKHDNLKSITVKAGATVRWEVKIGGEPIPEVKWFKGNQ 2132  
QY 106 Q-----Q-----A-----A-----A----- 110  
Db 2133 QLENGIQLTIDTRKNEHTILCIPSAMRSVDGEYRLTVKNSHGADBEKANLTVLDRPSKN 2192  
QY 111 -----A-----A-----A-----A-----A----- 113  
Db 2193 GPLEVSDVPEDNLNLSWKPPDDGGEPIEYVEVEKLDATGRWVPCAKVKDTKAHIDGLK 2252  
QY 114 -----A-----A-----A-----A-----A----- 118  
Db 2253 KGQYQFRVKAVNKEGASDALSTDKDTAKNPNYDEFGKTGTGTPDWDADRVSLWEPPK 2312  
QY 119 S-----Q-----Q-G-S-----A-----K-N-----GE 128  
Db 2313 SDGAPITQYVIEKKGHGHDWQECGVSGDQTNAEILGLKEGEYQFRVAVNKAAGPE 2372  
QY 129 -----N-----T-----AN-----GE-----E-N----- 136  
Db 2373 ASDPSRKVAKPNLKPWIDREAMKTTIKVGNDFVEFDVVRGEPPEPKKEWIFNEKPVDD 2432  
QY 137 -----GA-----H-----TI-A-N----- 143  
Db 2433 QKTRIESDYKTRVLRGATRKHAGLYTLTATNASGSDKHSVEVIVLGRPSSPLGLEVS 2492

QY 144 N-----H-----TDM-----ME----- 150  
Db 2493 NVYEDRADLEWKVPEDDGGAPIDHYIEKMDLATGRWVPCGRSETTKTTVNLNLPQGHYK 2552  
QY 151 -----V-----D-----G-----D-----V 155  
Db 2553 FRVAVNKEGESDPLTNTAILAKNPEYVPGVKDPBELVDWDKXHDVLDLAWNAPDDGAPI 2612  
QY 156 E-----I-----P-----NKA-----V 163  
Db 2613 EAFVIEKKDKNGRWEALVVPDQKTATVNLKEGEYQFRSARNKAGTGDPSPDRV 2672  
QY 164 V-----J-----R-----G-----H-----ES----- 170  
Db 2673 VAKPRNLAPRIHREDLSDTTTRKVGATLKFIVHIDGEPAPDVTWSPNGKIGESKAQIENE 2732  
QY 171 ----- 172  
Db 2733 PYISRFPALPKALRKQSGKYTITATNINGDSVTINIKVSKPTKPKGPIEVTDFEDRAT 2792  
QY 173 -----F-----I-----C----- 175  
Db 2793 LDWKPPEDDGGEPFIEFYIEKMTKDIWVPCGRSGDTHFTVDSLNGDHYKFRVKA VNS 2852  
QY 176 ----- 176  
Db 2853 EGSPDPLETETDILAKNPFDRDRPGRPEPTDWDSDHVDLKWDPPLSDGAPIEYQIEK 2912  
QY 177 -----W----- 177  
Db 2913 RTKYGRWEPAITVPGQTTATVPDLTPNEEYFRVAVNKGSPSPDASKAVIAKPRNL 2972  
QY 178 -----N-----PVS-----D-----L----- 183  
Db 2973 KPHIDRALKNLTIKAGQISFDVPVSGEPAPTVTWHWPDNREIRNGRVLKDNPEYQSK 3032  
QY 184 L-----A-----S-G-----S-GD----- 190  
Db 3033 LVVKQMERGDSGTFTIKAVNGEDEATVKINVDKPTSPNGPLDVSVDVHGHVTLNWRA 3092  
QY 191 -----S-----TA-----R----- 194  
Db 3093 PDDGGIPIENVVIEKYDTASGRWVPAKAVAGDKTTAVVDGLIPGHEYKFRVAAVNAEGE 3152  
QY 195 ----- 195  
Db 3153 SDPLETFGTTLAKDPFKGTNAPEITDWDKXHDVLEWKPPANDGGAPIEYVVEWKDE 3212  
QY 196 -----W-----NL-----S-E----- 200  
Db 3213 FSPFNDVAHVAGQTNATVGNLKEGSKYEFIRAKNKAAGLGDPSDSASAVAKARNVPPV 3272  
QY 201 -----NS-----T-----S-----GST----- 207  
Db 3273 IDRNSIQEIKVKAGQDFSLNIPVSGEPTTITWTTEGTPEVSDDRMKNLNEGKTKFHVK 3332  
QY 208 Q-L-----VL-----R----- 212  
Db 3333 RALRSDTGTYYIIKAENENGTDTAEVKVTVLDHPSSPRGLDVTNIVKDCGLAWKEPEDD 3392  
QY 213 -----H-----C----- 214  
Db 3393 GGAESHVYIEKQDAATGRWTACGESKOTNFHVDDLTOGHEYKFRVAVNRHGDSDPLEA 3452  
QY 215 -----I-----R-----EG----- 218  
Db 3453 REAIIAKDPFDRADKPGTPEIVDWKDHADLKWTPPADDDGAPIEGYLVEMRTPSGDWVP 3512  
QY 219 ----- 218  
Db 3513 AVTVGAGELTATVDGLKPGQTYQFRVKALNKAGESTPSPSRTMVAKPRHLAKINRDMF 3572



QY 219 -----GQ-----DV-----P----- 223  
Db 3573 VAQVRKAGQTLNFDVNVGEPAKPIBWFLNGSLSSGNGTHIDNNTDNNTKLTKKSTARA 3632  
QY 224 -----S-----N-----KQVT--S-----LD----- 232  
Db 3633 DSGKYIVATNESGKDEHEVDNILDIPGAPEGLRHKDITKESVVLKWDPELDGSGSPI 3692  
QY 233 -----W-----N-----SE-----GT--L----- 239  
Db 3693 TNYVVEKQEDGRWPCGETSDTSLKVNKLSEGEHYKFRKAVNRQGTSAPLTSDHAIYA 3752  
QY 240 -----L-----A-----T-----G-----SY-----DG 247  
Db 3753 KNPFDEPDAPDVTVPDWDKOHVLEWKPANDGGAPIDAYIVERKDKFGDMWECARVDG 3812  
QY 248 -----F-----F-----A-----RI-- 251  
Db 3813 KTTKATADNLTPGETTYQFRKAVKAVKAGPKGKPSDPTGNVAVKPRMAPKLNLAGLLDLRIK 3872  
QY 252 -----W-----TKD--G-----NL-----A-----S-----T--L-- 262  
Db 3873 AGTPIKLDIAFEGEPAPVAKWKANDATIDTGARADVTNTPTSSAIHIFSAVRGDTGVYKI 3932  
QY 263 -----G-----Q-----HK--G-----P-----I-- 269  
Db 3933 IVENEHGKDTAQCNVTVDVPGTPPEGLKIDIEHKEGCTLNWKPPDNGGTDVLHIVEK 3992  
QY 270 -----F-----A-----L----- 272  
Db 3993 MDTSRGTWQEVGTFPDCTAKVNKLVPGEYAFRKA VNLQGESKPLEABEPIIAKNQFDV 4052  
QY 273 -----KWN-----G-----G----- 278  
Db 4053 PPVDKPEVTDWKDIDIKWNPNTANGAPVPTYGYIVERKKGSAIWEAGKTPGTTTFA 4112  
QY 279 -N-----F-----IL--S-----AG----- 285  
Db 4113 DNILKPGVEYFRVIAVNAAGSPSPDPTDQITKARYLKPILITASRKIKIKAGFTHNLE 4172  
QY 286 -----VD--K--TT----- 290  
Db 4173 VDFIGAPDPTATWVGDSGAALPELLVDAKSSTTTIFPFSAKRADSGNYKLVKNELGE 4232  
QY 291 -----I-----I-----W-----D----- 294  
Db 4233 DEAFIEVIQDRPSAPEGPLEVSDVTKDSCLVNLWKPDKDGGABISNYVVERKDKTKNTW 4292  
QY 295 -----A-----H-----T--G-----E-----AKQOF----- 304  
Db 4293 VPVSASFVTGTSITVPKLTGEGHEYFRVMAENTFGRSDSLNTDPEVLAKDPFGTTPGKGRP 4352  
QY 305 -----P-----P-----P----- 306  
Db 4353 EIVDTDNHDIDIKWDPDRNGSGPVHDYDIERKDAKTGRWKVNTSPVQGTAFSDTRVOK 4412  
QY 307 -H-----SA-----P-----A-----LD-----V----- 314  
Db 4413 GHTYFRVAVNAKAGQPSDSAAATAPKMPHEAPKFDLDLGGKFRVKAPELVTITPF 4472  
QY 315 -----D--W-----Q-----S----- 318  
Db 4473 TASPDQISWTEGKGLACVETDTSQTKLVIPSTRSDSGPVKIKAVNPYGEAEANIKI 4532  
QY 319 -----N-----N----- 320  
Db 4533 TVIDKPGAPENITYPAVSRHTCTLNWDAPKDDGAGIAGYKIEYQVSGQIWDKVPGLIS 4592  
QY 321 -----T-----F-----A-----S--C-----ST----- 327  
Db 4593 GTAYTVRGLEHQOYFRIRAENAVGLSDYCOQVPVVIKDPDPGAPSTPEITGYDTNQ 4652  
QY 328 -----D-----M-----CI----- 331

Db 4653 VSLAMNPPRDDGSGSPILGYVVERFEKRGGDWAPVKNPMVKTECIVPGLHENETYQFRV 4712  
QY 332 -----H-----V--C----- 334  
Db 4713 RAVNAAGHEPNSGSEBVTCTRPYVEKPGADAPRVGKITKNSAELTWNRLRDLRGGAIDG 4772  
QY 335 -----KLG----- 337  
Db 4773 YIVEKKLGDNDWTRCNDKPVDRDTAFEVKNLGEKEEYFRVIAVNSAGEBEPKSPDLVL 4832  
QY 338 -----Q-----D-----R-----PI----- 342  
Db 4833 IBEQGRPIFDINNLKIDITVRAGETIIRIPYAGGNPKPIIDLFGNSPIFENERTVVDV 4892  
QY 343 -----K--T-----F-----Q----- 346  
Db 4893 NPGSEIVITTTGSKRSDAGPYKISATNKYKOTCKLVFLDAPGKTGPIRATDIQADAM 4952  
QY 347 -----GH-----T-----N--B--V--N-- 353  
Db 4953 TLSWRPPKONGDAITNYVVEKRTPGGDWTVVGHVPGTTLRVNLDANTPYEYFRVRAEQ 5012  
QY 354 -----AI--K-----W-----D--P----- 359  
Db 5013 YGVGEPLETDDAIVAKNPPFTPGAPQPEAVETSEBAITLQWTRPTSDGAPIQGVYIEK 5072  
QY 360 -----T-----GN----- 362  
Db 5073 REVGSTWTKAAGNILDTKHRVTGLTPKTYEYFRVAAVNAAGQGSYSVNSVPITADNAP 5132  
QY 363 -----L----- 363  
Db 5133 TRPKNMGMLTRDILAYAGERAKILVPPAASPAPKVTFSKGENKISPTDPRVKVEYSDFL 5192  
QY 364 -----L-----AS-----C----- 367  
Db 5193 ATLTIKSELTDGGLYFVELENSQSDSASIRLVKVDKPSAQHIRVEDIAPDCTLYWM 5252  
QY 368 ---SD-----DM-----T-- 372  
Db 5253 PPSDGSQSPITNYIVEKLDLRHSDGKWKVSVFVRNLNTVVGGLIKDNRFRFRVRAETQY 5312  
QY 373 -----L-----K-----I-----W-----SMK-----Q-- 380  
Db 5313 GVSEPCELADVVAKYQVEVPNQPEAPTVRDKDSTWAELEWPPRGGGSKIIQYQVQYR 5371  
QY 381 D-----N-----CVHD-----L--Q-----Q----- 389  
Db 5372 DTSSGRWINAKMDLSEQC--HARVTGLRQNGEPEFRIIAKNAAGFSKPSPPSERCOLKSRF 5430  
QY 390 -----H-----N-----K-----EI--Y--TIKW--S----- 400  
Db 5431 GPPGPIHVGAKSIGRHNHCTITWMAPLEDGGSKITGYNVEIREYGSTL--WTVASDYNVRE 5489  
QY 401 P--T-----GP-----G-----T-----N 407  
Db 5490 PEFTVDKLEFNDYFRVVAINAAGKIPSLPSGPIKIOESGSRPOIVVVPEDTAQPN 5549  
QY 408 -----N-----P----- 409  
Db 5550 RRAVFTCEAVGRPEPTARWLNRNGRELPESSRYRFEASDGVYKFTIKEVMDIDAGEVTV 5609  
QY 410 -N-----ANL----- 413  
Db 5610 SNPYGSDTATANLVQAPPVIBKDVNTILPSGDLVRLKIYFSGTAPFRHSLVLRNREID 5669  
QY 414 M-----L-----A--S--A-----F----- 420  
Db 5670 MDHPTIRIVEFDHILITIPALSVREAGRYEYTVVNSDSEANTGFWLNTGLPEAQGPL 5729  
QY 421 -----D--S-----TV----- 424

Db 5730 HISNIGPSTATLSWRPPVTDGSKITSYVVEKRDLSKDEWVTVTSNKMNYIIVTGLFEN 5789  
QY 425 ----- 424  
Db 5790 HEYFRVSAQNEGIGAPLVSEHPHIIARLPDPPTSPNLNLEIVQGGDYVTLISWQRPUSD 5849  
QY 425 --BL-----W----- 427  
Db 5850 GGGRLRGYIVKQBEHDEWFRCNQNPSPNNYVNPNLIDGRKYRYRVFAVNDAGLSOLA 5909  
QY 428 --D-----V---D-----R----- 431  
Db 5910 ELDTQLFOAGSGGPKIVSPLSDLNNEVGRCVTFECEISGSPREYRWFKGCKELVDTS 5969  
QY 432 ---G-----I-----C-----I---H-- 436  
Db 5970 KYTLINKGDKQVLIINDLTSDADDEYTCRATNSSGSTRANLRIKTKPRVFPPIPKYHGG 6029  
QY 437 -----T-----TL----- 438  
Db 6030 YEAQGETIELKIPYKAYPQGEARWTKDGEKIENNSKFSITDDKFAILRISNASREDYG 6089  
QY 439 -----S-----TL----- 438  
Db 6090 EYRVVENSVDGSGTVNVTADVPEPPRPPIENILDEAVILSKPPALDGGSLVTNYT 6149  
QY 439 -----T-----KH-Q-----E---PV----- 445  
Db 6150 IEKREMGGSWSPCAKSRVYTTTIEGLRAGQYEFRIIAENKKGQSKPCEFTAPVLIPGD 6209  
QY 446 -----Y-----S-----V----- 448  
Db 6210 ERKRRGYDVEQKIVRGKGTSSNYDNYVDIWKQYYPQVVEIKHDHVDHYDHEEL 6269  
QY 449 ---AF-----S-----P----- 452  
Db 6270 GTGAFGVHVRTERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLVNLHDAFED 6329  
QY 453 D-----G-----R-----Y----- 456  
Db 6330 DNEMWMIYFWSGGLFEKVADEHNKMSDEAVEYMRQVCKGLCHMHNHYVHLDLPEN 6389  
QY 457 -----L-----A----- 458  
Db 6390 IMFTKRSNELKLIDFGLTAHLPKQSVKVTGTGAFAAPEVAEGKPVGYVTDWMSVGL 6449  
QY 459 ---SG-S-F---D---K-C-----V 466  
Db 6450 SYLLSGLSPFGGENDDETILRNKSCDMNMDSAFSGISBDGKDFIRKLLADPNTRMTI 6509  
QY 467 H-----I-----W-----N----- 470  
Db 6510 HQALEHFWLTPGNAPGRDSQIPSSRYTKIRDSIKTKYDAMPEPLPLGRISNYSYLRKHR 6569  
QY 471 -----TQ-----T-----G-A-L-----V-H----- 478  
Db 6570 PQEYSIRDAFWRSEAPREFIVKPYGTGVEGOSANFYCRVVIASSPPVVTWHKDDRELKQ 6629  
QY 479 S--Y--R-----G-T-----G-----I-----FE--- 488  
Db 6630 SVKTMKRYNGNDYGLTINRVKDDKGEYTVRAKNSYGTKEEIVFLNVRHSEPLKPEPLE 6689  
QY 489 -----V-----C-----W----- 491  
Db 6690 PMKAPSPRVEBFKERSAPFFTHLRNPLIOKNHQCKLTCSLQGNPNPTIEWKDGHP 6749  
QY 492 -----NA-----AG-----D-----K-----V 498  
Db 6750 VDEDRVQVSPRSGVCSLEIFNARVDDAGTVTATNDLGDVDSVCEVLTVQTKGGEPIPV 6809  
QY 499 -----G-----A-----SA----- 502  
Db 6810 SSFPRRAYDTLSTGTVDVERSHYADMRRLSLRDSVPDRSAADLTKITNELPSFTA 6869

QY 503 --SD-----GS-----V----- 507  
Db 6870 QLSDSSETVEVGSAAEFAAVSAGQPELIEWLHNGERISESDSRFRASVYAGKATLRISDAK 6929  
QY 508 -----C-----VL-----D-----LR--K 514  
Db 6930 KSDEGYLCRASNSAGOEQTRATLTGVKQDPLLNGHAGQAVESELRTVK 6978

RESULT 13  
T12117

polyprotein - fava bean dsRNA replicon

C:Species: Vicia faba (fava bean)

C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004

C:Accession: T12117

R:Pfeiffer, P.

J. Gen. Virol. 79, 2349-2358, 1998

A:Title: Nucleotide sequence, Genetic organization and expression strategy of the double-

A:Reference number: Z17424; MUID:98451319; PMD:9780039

A:Accession: T12117

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-5825 &lt;PFE&gt;

A:Cross-references: UNIPROT:O82731; EMBL:AJ000929; NID:g3184155; PIDN:CAA04392.1; PID:g31

A:Experimental source: virion; cultivar 447

C:Comment: This gene product may be cleaved into several proteins including helicase and

C:Genetics:

A:Genome: dsRNA replicon

C:Superfamily: fava bean dsRNA replicon polyprotein

Query Match 70.3%; Score 2571.1; DB 2; Length 5825;  
Best Local Similarity 9.1%; Pred. No. 3.9e-52;  
Matches 442; Conservative 56; Mismatches 13; Indels 4349; Gaps 376;

QY 1 M---SIS-----S-----D-----E---V-----N-- 9  
Db 949 MATRPSITQDAPVMKESNTLLIDMNLRLMDPMWLITECLRRKITTAYVTPMLELDQNP 1008  
QY 10 -----F-----LV-----Y----- 13  
Db 1009 QTILGEVQFRKVENQIVITAWSORPIRVTTENYMNWTKVDAFVIGENTYPTDVTREAN 1068  
QY 14 ---R-----SG-----PS-----HS-AF-- 26  
Db 1069 GISLRIDLAASTFEMSEVETVGDSEISVPTVYVYKPKRKLMSGEKKVKULREL 1128  
QY 19 -----SG-----PS-----HS-AF-- 26  
Db 1129 YRRMMIRNVSGTSFKELVDYIGISNMKNYNINDMVVSFSDISDDTIRHTAAAFVYSKMIT 1188  
QY 27 -----T---F----- 28  
Db 1189 RANSMLQTERVLESVNTARGFLASLLISOAETMLKITGLDETMLTTVNRWMSKGEVRL 1248  
QY 29 -----G-----K----- 31  
Db 1249 KGMNLNLSLIDNLNLCYSREPVIHRLRAQTELEPTGTCPHKTPDFVKECSMCECG 1308  
QY 32 -----S-----H----- 33  
Db 1309 VNPATGAIGFCSCSLPTDKHQCDHPCKHAHESEAGTRCSCCLPIAGEACPGVNRQIE 1368  
QY 34 --I-----S-----Q-----SNIN--- 40  
Db 1369 SEILFENSESEADEAEQNRVRKNSKPRPRDREKGNRNNRNNDSRRRTDANHHSNVNHH 1428  
QY 41 -----GALV--PP--AAL-----I---S-----I 52  
Db 1429 GHNKQRQGA--TQQPPKSAALHPENDNDTDPSTFTVLTADPTNPPGEPQAGDNI 1487  
QY 53 -----I-----Q-----KG-----L-Q-----Y---V 60



QY 319 -N---N---T---F---AS---C---STD--- 328  
 Db 3706 QAIIVCQLLGLNCCLIQNGETGTVNFAPNPFVQLMRLSODASYDHCVLINLLGADGVK 3765  
 QY 329 -M---C---I---HV---C--- 334  
 Db 3766 RLSPENIAKENQLEHVCPVDNKPIVAVGENPYACVSHVIDNEDLARLTSCSQPYEKI 3825  
 QY 335 -K--- 335  
 Db 3826 PVLDRFEMLPVNNVMARMTKTSATLLKKVTPITTGAGSLALTRNYKATPMVTTKPG 3885  
 QY 336 -L---G---Q--- 338  
 Db 3886 DVLVLRDHCCKVMCTHTTIVDPGGVEQILEITDDIGDGAVALDILGLHCLTPADRTPR 3945  
 QY 339 -DRP---IKT---F---OG---H---T--- 349  
 Db 3946 RDRPVNTSNEVWSLNIETQRYLEATKGLRSHVGQEGKLIHYHFDNREHMYDEHYIDA 4005  
 QY 350 -N---E---VN---A---IK-W---D-P--- 359  
 Db 4006 MLDENIVIELDPTVQVPTGVNIVNLICIRRMFLAMLYGPTWVAYCCDNFSAQLAWFL 4065  
 QY 360 -A---S---CS---DD-M--- 371  
 Db 4186 DIKQCAVHVPTGKVALRVVVSLSHEIACSTKVYGVDPKDDMRAGASNLTSVPASDX 4245  
 QY 372 -T---L---L---KI 375  
 Db 4246 ADPGSSQLSCEENWSQTGPVACVHIEPCDHAANASNDVLKEITQPEQNHHIISKI 4305  
 QY 376 -W---S---M--- 378  
 Db 4306 LATAQAKASPEDKSMKSGATSADLQAIQSILCDSPGMSQVADQLAKLEVNFSDVT 4365  
 QY 379 -K---Q---D--- 381  
 Db 4366 LSGIHRORNVNSQGYLQVRDISERWAIYKGIHGHSDVPTNPNPGWMLAPFPVRKLG 4425  
 QY 382 -NC---V---H---D-L---Q--- 388  
 Db 4426 VYKEKEIKVTMNVAGNCAYRLNRIKREP KSVFTHNKPEVMDILITYTSANNKQYPTN 4485  
 QY 389 -Q---H---N---K-E-I---Y--- 395  
 Db 4486 LORLADKYLHGRVSHVYVYEGYNRDLQTCVELTKRMERVIWVTDAPDVYNQLDADIQV 4545  
 QY 396 -TK---W---SP---TG---P--- 404  
 Db 4546 REKIITIKVTSQPEPYRAIWRFWILPANIDPECELYILNGREWAPNDMELVDRVMEESP 4605  
 QY 405 -G---T---N---N--- 408  
 Db 4606 EDFDIVSALHCOAFNVCCGYAKYNHAKFTDYNTFNDIYEAGNSYGGDEIYMSRYKVKWM 4665  
 QY 409 -P-NA---NML---A---SASF--- 420  
 Db 4666 WLYYILPGGAWYGIEGKWLNALYDITICLEKEGERTRIDHVKGLFCSANFFHVPDITKG 4725  
 QY 421 -D-STV---R---L--- 426  
 Db 4726 IYDYSTVDMSVDGIPWQYAMQILINGHSVTENPVHSTGESTGNPRINSDGKXPWQLNPN 4785  
 QY 427 -W-D---V---D---R-G---I---C--- 434

Db 4786 NWLLDPGLKKSEKMKIQKFLDSSSVKCIKESOSLPQLGPKLPVVVFANQACEAIGVSM 4845  
 QY 435 -I---H-TL--- 438  
 Db 4846 NNTGTLVVEVNIQAQKLAARNEPLOHQTLILKSHLILMSTTGRHASKWPKPEYTHLMKADLQ 4905  
 QY 439 -T---K---H-Q---E---P---V--- 445  
 Db 4906 NCQVYVEGSATSILOWKRETOHAQVYTTTEERDVFRIPQALDHSVGTLLKLVKVTGNE 4965  
 QY 446 --YS---V---A---F---S---P--- 452  
 Db 4966 LGLSSSSGVLARVLKNDRLANNHLETOHLQKLEFDRRIISSHBESSGMTKPRIGQT 5025  
 QY 453 -D---G--- 455  
 Db 5026 LYLADVLTAAIGIECVNHITTSREWEKALKNPITKVGMPLTNSVLVSGKLGCVQPTIPR 5085  
 QY 456 -Y---L--- 459  
 Db 5086 CYAMLCQGVKLIKFPETRGGTTPSNVSANVGSNPTTAGPGDTPPLITSWNKVESGAS 5145  
 QY 460 -G-S---FD---K--- 464  
 Db 5146 NQSPFGVSASAKAYQDKLREKKIETPQDRLRVLETMFLDDBAYALMSHGLTDLRKLNDGGNWH 5205  
 QY 465 C---V---H---I---W--- 469  
 Db 5206 CATDATMKSDFEIIEPVGGDLPLSLNTSALRHNETNARVIDLWDDTDLRDLWLTLYAPKNP 5265  
 QY 470 -N---TO---TG--- 474  
 Db 5266 MKLTSRVSPQGGKINILKTLTNRPCQTPVPTQVMGNGENAVTGLRGLSVLPLRREPNNVT 5325  
 QY 475 -AL--- 476  
 Db 5326 HELHKPTAYRDGWERVLDKFKANTITISDADVKTWLSRRSDWKALATSTIKWLTGLP 5385  
 QY 477 -VH--- 478  
 Db 5386 SNPMAVNVHVHTESLLKANPIMYWRQTQGRIVMQPKELCALMSPAFIAIKRRLKEVLR 5445  
 QY 479 -S---Y--- 480  
 Db 5446 DEIVYTDGLTPDMLSRARTIQVDYVVFEDDLVIQDRQTQDELIDLEFQWMLDLGLDLINLA 5505  
 QY 481 -R---G---TG--- 484  
 Db 5506 NLWRLVHNKWRPKGQHSWGQDAMRLTGOATTALGNATNLVHSSSFVIEHROAIKLMFV 5565  
 QY 485 -G-I-F---E---V---C--- 490  
 Db 5566 LGDDNITFMSABPNLTQYKRLMSERYNMRKSKPQVSRNVGTFCSLLCYRNSFGHCEVGPDP 5625  
 QY 491 -W-N- 492  
 Db 5626 VRLRHRFEVTGGNHYELPVAVKMPHNVRIGKTFPNGRYLHVMEENEEAEFVEWLRNH 5685  
 QY 493 A---G---D-K---VGAS---A--- 502  
 Db 5686 ANRSYILVLKTSNAEWSRLDGVQVHDLSKIDTSGNLAEALGIGASMGMPNPMMAAGMSY 5745  
 QY 503 -S-D---G---S-VCV---L-D---L-RK 514  
 Db 5746 LHMGLKSEADKINTEKEFGLRKYWDVASCNATCIKYNMSPMELEDNIGELLMIRK 5805

RESULT 14

S25111

alpha-2-macroglobulin receptor precursor - mouse

N;Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor  
 C;Species: Mus musculus (house mouse)

C>Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 09-Jul-2004  
C/Accession: S25111; S32554  
R:van Leuven, F.  
submitted to the EMBL Data Library, July 1992  
A:Reference number: S25111  
A/Accession: S25111  
A/Molecule type: mRNA  
A/Residues: 1-4545 <VAN1>  
A/Cross-references: UNIPROT:Q61291; EMBL:X67469; NID:G49941; PIDN:CAA47817.1; PID:G49942  
R:van Leuven, F.; Stas, L.; Raymakers, L.; Overbergh, L.; de Strooper, B.; Hilliker, C.;  
Biochim. Biophys. Acta 1173, 71-74, 1993  
A/Title: Molecular cloning and sequencing of the murine alpha-2-macroglobulin receptor c  
A/Reference number: S32554; MUID:93250049; PMID:8485155  
A/Accession: S32554  
A/Status: nucleic acid sequence not shown  
A/Molecule type: mRNA  
A/Residues: 1-281416-4453 <VAN2>  
A/Cross-references: EMBL:X67469  
C:Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated  
clated protein (see PIR:JX0281).  
C:Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding  
C:Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; glycopro  
P:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-3944/Domain: Product: alpha-2-macroglobulin receptor #status predicted <NAT>  
F:20-3944/Domain: alpha-2-macroglobulin receptor 515K chain #status predicted <S15K>  
F:28-65/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F:73-109/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F:116-149/Domain: EGF homology <EG1>  
F:155-189/Domain: EGF homology <EG2>  
F:199-240/Domain: LDL receptor YWTD-containing repeat homology <YW01>  
F:241-282/Domain: LDL receptor YWTD-containing repeat homology <YW02>  
F:293-335/Domain: LDL receptor YWTD-containing repeat homology <YW03>  
F:336-379/Domain: LDL receptor YWTD-containing repeat homology <YW04>  
F:380-421/Domain: LDL receptor YWTD-containing repeat homology <YW05>  
F:422-469/Domain: LDL receptor YWTD-containing repeat homology <YW06>  
F:479-520/Domain: EGF homology <EG3>  
F:515-560/Domain: LDL receptor YWTD-containing repeat homology <YW07>  
F:615-660/Domain: LDL receptor YWTD-containing repeat homology <YW08>  
F:661-711/Domain: LDL receptor YWTD-containing repeat homology <YW09>  
F:712-753/Domain: LDL receptor YWTD-containing repeat homology <YW10>  
F:754-800/Domain: LDL receptor YWTD-containing repeat homology <YW11>  
F:808-843/Domain: EGF homology <EG4>  
F:855-891/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
F:896-932/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
F:937-972/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
F:977-1012/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
F:1016-1052/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
F:1063-1098/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F:1105-1141/Domain: LDL receptor ligand-binding repeat homology <LDL9>  
F:1146-1183/Domain: LDL receptor ligand-binding repeat homology <LDL9>  
F:1186-1222/Domain: EGF homology <EG5>  
F:1228-1262/Domain: EGF homology <EG6>  
F:1270-1309/Domain: LDL receptor YWTD-containing repeat homology <YW12>  
F:1310-1356/Domain: LDL receptor YWTD-containing repeat homology <YW13>  
F:1357-1399/Domain: LDL receptor YWTD-containing repeat homology <YW14>  
F:1400-1446/Domain: LDL receptor YWTD-containing repeat homology <YW15>  
F:1447-1489/Domain: LDL receptor YWTD-containing repeat homology <YW16>  
F:1490-1532/Domain: LDL receptor YWTD-containing repeat homology <YW17>  
F:1541-1579/Domain: EGF homology <EG7>  
F:1584-1627/Domain: LDL receptor YWTD-containing repeat homology <YW18>  
F:1628-1670/Domain: LDL receptor YWTD-containing repeat homology <YW19>  
F:1671-1714/Domain: LDL receptor YWTD-containing repeat homology <YW20>  
F:1715-1754/Domain: LDL receptor YWTD-containing repeat homology <YW21>  
F:1758-1797/Domain: LDL receptor YWTD-containing repeat homology <YW22>  
F:1798-1847/Domain: LDL receptor YWTD-containing repeat homology <YW23>  
F:1851-1887/Domain: EGF homology <EG8>  
F:1935-1977/Domain: LDL receptor YWTD-containing repeat homology <YW24>  
F:1978-2020/Domain: LDL receptor YWTD-containing repeat homology <YW25>  
F:2021-2064/Domain: LDL receptor YWTD-containing repeat homology <YW26>  
F:2065-2106/Domain: LDL receptor YWTD-containing repeat homology <YW27>  
F:2107-2152/Domain: LDL receptor YWTD-containing repeat homology <YW28>  
F:2160-2195/Domain: EGF homology <EG9>  
F:2200-2242/Domain: LDL receptor YWTD-containing repeat homology <YW29>

F:2254-2295/Domain: LDL receptor YWTD-containing repeat homology <YW30>  
F:2345-2389/Domain: LDL receptor YWTD-containing repeat homology <YW31>  
F:2390-2430/Domain: LDL receptor YWTD-containing repeat homology <YW32>  
F:2431-2474/Domain: LDL receptor YWTD-containing repeat homology <YW33>  
F:2483-2518/Domain: EGF homology <EG10>  
F:2525-2562/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F:2567-2601/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F:2606-2640/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F:2653-2689/Domain: LDL receptor ligand-binding repeat homology <LDLE>  
F:2697-2731/Domain: LDL receptor ligand-binding repeat homology <LDLF>  
F:2735-2770/Domain: LDL receptor ligand-binding repeat homology <LDLG>  
F:2775-2813/Domain: LDL receptor ligand-binding repeat homology <LDLH>  
F:2819-2854/Domain: LDL receptor ligand-binding repeat homology <LDLI>  
F:2859-2898/Domain: LDL receptor ligand-binding repeat homology <LDLJ>  
F:2905-2940/Domain: LDL receptor ligand-binding repeat homology <LDLJ>  
F:2945-2981/Domain: EGF homology <EG11>  
F:2987-3022/Domain: EGF homology <EG12>  
F:3030-3069/Domain: LDL receptor YWTD-containing repeat homology <YW34>  
F:3070-3114/Domain: LDL receptor YWTD-containing repeat homology <YW35>  
F:3115-3157/Domain: LDL receptor YWTD-containing repeat homology <YW36>  
F:3158-3201/Domain: LDL receptor YWTD-containing repeat homology <YW37>  
F:3202-3242/Domain: LDL receptor YWTD-containing repeat homology <YW38>  
F:3243-3285/Domain: LDL receptor YWTD-containing repeat homology <YW39>  
F:3295-3331/Domain: EGF homology <EG13>  
F:3335-3370/Domain: LDL receptor ligand-binding repeat homology <LDLL>  
F:3375-3409/Domain: LDL receptor ligand-binding repeat homology <LDLM>  
F:3414-3449/Domain: LDL receptor ligand-binding repeat homology <LDLN>  
F:3454-3490/Domain: LDL receptor ligand-binding repeat homology <LDLO>  
F:3493-3532/Domain: LDL receptor ligand-binding repeat homology <LDLP>  
F:3537-3571/Domain: LDL receptor ligand-binding repeat homology <LDLQ>  
F:3576-3610/Domain: LDL receptor ligand-binding repeat homology <LDLR>  
F:3614-3648/Domain: LDL receptor ligand-binding repeat homology <LDLS>  
F:3655-3691/Domain: LDL receptor ligand-binding repeat homology <LDLT>  
F:3698-3732/Domain: LDL receptor ligand-binding repeat homology <LDLU>  
F:3742-3777/Domain: LDL receptor ligand-binding repeat homology <LDLV>  
F:3786-3823/Domain: EGF homology <EG14>  
F:3829-3861/Domain: EGF homology <EG15>  
F:3869-3912/Domain: LDL receptor YWTD-containing repeat homology <YW40>  
F:3913-3970/Domain: LDL receptor YWTD-containing repeat homology <YW41>  
F:3945-4345/Domain: alpha-2-macroglobulin receptor 85K chain #status predicted <85K>  
F:3945-4421/Domain: 85K chain extracellular #status predicted <85K>  
F:3971-4013/Domain: LDL receptor YWTD-containing repeat homology <YW42>  
F:4014-4057/Domain: LDL receptor YWTD-containing repeat homology <YW43>  
F:4058-4100/Domain: LDL receptor YWTD-containing repeat homology <YW44>  
F:4101-4143/Domain: LDL receptor YWTD-containing repeat homology <YW45>  
F:4152-4183/Domain: EGF homology <EG16>  
F:4201-4232/Domain: EGF homology <EG17>  
F:4237-4268/Domain: EGF homology <EG18>  
F:4273-4304/Domain: EGF homology <EG19>  
F:4309-4340/Domain: EGF homology <EG20>  
F:4345-4375/Domain: EGF homology <EG21>  
F:4378-4409/Domain: EGF homology <EG22>  
F:4422-4445/Domain: Transmembrane #status predicted <TM>  
F:4446-4545/Domain: Intracellular #status predicted <INT>  
F:167,2999/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F:2953/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted  
F:4076,4126,4279/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match

Best Local Similarity 70.2%; Score 2569.5; DB 1; Length 4545;

Matches 420; Conservative 69; Mismatches 17; Indels 3807; Gaps 358;

QY 1 MS-----I-----SSDE-----V-----N-F-I-----11

DB 89 MSRLNGIQDCWMDSGEAGHCRANCRMCQHCVPPTSGPTCYCNSSFOLEADGKT 148

QY 12 -----VY-----R-----Y-LQ-----E-----18

DB 149 CKDFECVSYGTCQLCTNDGSGFTCCGCGVEGVLQPDNRCKAKNEPVDPPVILLANSQ 208

QY 19 -----SG-----PS-----H-----SA-----F 26

DB 209 NILATYLSGAQVSTITPTSTRTTAMDFSYANETCVWHVGDSSAAQTQLKCARMPGLKGF 268

QY 27 ---TFGIK-S-HISQ---S-N---I---NG--- 41  
Db 269 VDEHTIINISLHHEVQWALDWLTGNFYFVDDIDRIFVQCNRGDTCVTLLDLELYNPKG 328  
QY 42 -AL-----VP-----P-----A-A- 48  
Db 329 IALDPANGKVFFTDYQIPKVERCDMGQNRKLVDSKIVFPHGITLIDLSRLVYWADAY 388  
QY 49 -LI-----SIIQK-L-Q--Y--V-E-----A-----EVS-I---N 67  
Db 389 LDYIEVDVYEGKGRQTIQ-GILIEHLYGLTVFENLYATNSDNANTQOKTSIVRNRFN 447  
QY 68 -E-----D-GTL----- 72  
Db 448 STEYQVTRVDKGGALHIYHQRQPRVRSACENDQYKPGGCSIDICLLANSHKARTCR 507  
QY 73 -F-----DG-----R-----PIE-----SL-- 81  
Db 508 RSGFSLGSGSKCKPEHELFLVYKGRPGIIRGMDMGAKVPDEHMIPIENLNPALDF 567  
QY 82 -----S-LI-----D-----AV--M-----PD-- 90  
Db 568 HAETGFYFADTTSYLIQKIDGTERETILKQDGIHNVEGVAVDMGDNLYWTDDGPKT 627  
QY 91 V-V-----QTR----- 97  
Db 628 ISVARLEKAAQTRKTLIEGKMTHPRAIVDPLNGMYMTDWEEDPKDSRRGLERAWMDG 687  
QY 98 -----A-Y-R-----D-K-----A-QQ----- 102  
Db 688 SHRDIFVTSKVLWPNGLSLDIPAGRLYWDAFYDRIETILLNGTDRKIVYEGPELNHAF 747  
QY 103 -----O-A-A-----L-----A-A-QQ----- 106  
Db 748 GLCHGNYLFWTEYRSGSVYRLERGAVAGPTVTLRSERPPIFEIRMYDAHEQQVGTNK 807  
QY 107 -----O-A-A-----A-----A-A----- 112  
Db 808 CRVNGCSSLCLATPGSRQCAEDQVLTDTGVTCLANPSYVPPQCGQGFACANNRC 867  
QY 113 -----A-A----- 114  
Db 868 IQERWKDGDNDCLNDSDEAPALCHOHTCPDRFKCENNRCPINRWLCGDNDGCGNSE 927  
QY 115 -A-A-A-----A-----AS-----Q----- 120  
Db 928 SNATCSARTCPNQFCASGRCIPISWTCDLDDCGDRSDESASCAYPCTFPLTQFTCNN 987  
QY 121 -----O-----G-----S----- 123  
Db 988 GRCININWCDNDGDSDEAGCSHSCSSTQFKCNSGRCIPEHWTCGDNDGCGYSDE 1047  
QY 124 -----A-----K----- 125  
Db 1048 THANCTNOATRPPGGCHSDEFQCLDGLCIPLRWCDGDTDCDMSDEKSCGCVTHVCDP 1107  
QY 126 N-----G-----EN-----T-----AN----- 132  
Db 1108 NVKFGKDSARCISKAWVCDGSDCEDNDEENCEALACRPPSHPCANNTSVCLPPDKLC 1167  
QY 133 -G-----E-E-----NG-----A-----HT--I--- 141  
Db 1168 DGKDDCGDSEBELCDQCSLNGGCHNCNVAPGEGIVCSPLGMELGSDNHTCQIQSY 1227  
QY 142 -A----- 143  
Db 1228 CAKHLKCSQKQKFSVKSCYEGWVLEPDGTCRSLDPFKLFIIFSNRHEIRRIDLHK 1287  
QY 144 -----N-----H-----TD-----MM----- 149  
Db 1288 GDYSVLVPLGRLNTIALDFHLSQSALYWTDAVEDKIYRGKLLDNGALTSFEVVIQYGLATP 1347

QY 150 -----EV--D-----GDVE---I---P----- 158  
Db 1348 EGLAVDWIAGNIYVWESNLDQIEVAKLDGTLRFTLLAGDIEHPRAIADPRDGLFTWDW 1407  
QY 159 ---P-----N-----K-----AV----- 163  
Db 1408 DASLPRIEASMSGAGRTIHRGTSGGCANGLTVDYLEKRIILWIDARSDAIYSARYDGS 1467  
QY 164 ---VLRGHE--S-----EV-----F--- 173  
Db 1468 GHMEVLRGHEFLSHPPFAVTLYCGEVYTDWRTWTLAKANKWTGHNVTVQRTWTQPDFLQ 1527  
QY 174 I-----C-A-----W-----N----- 178  
Db 1528 VYHPSQPMAPNCPCEANGRGPCSHLCLINARTVSWACPHMLKHKDNTTCYEFKKFL 1587  
QY 179 ---P-----VS-----D-----L-L-A-----S----- 186  
Db 1588 YARQMEIRGVLDLADAPYNYIISFTVPDIDNVTVLDYDAREQRYVWSDVTOAIKRAFING 1647  
QY 187 -G-----S-----G--D-----S--T-----AR----- 194  
Db 1648 TGVETVVSADLPNAHGLAVDWVSRNLFMTSYDTNKKQINVARLDGSKNAVQGLEQPHG 1707  
QY 195 --I-----W-----NLS-EN--S--T-----S----- 204  
Db 1708 LVVHPLRGKLYWTDGDNISMANNDGSHLLFSGQKGPVGLAIDFPESKLYWISSGNHTI 1767  
QY 205 ---GS-----T--QL-----C-I-----R----- 216  
Db 1768 NRCNLGDSLEVIDTWRSQKATALAIMGDKLWADQVSEKMGTCNKGADGSGSVLRNS 1827  
QY 213 -----H-----C-I-----R----- 216  
Db 1828 TTLVHMVKVYDESIOLEHEGTNPCSVNNGDCSQCLPTSETTRSCMCTAGYSLSRQOAC 1887  
QY 217 EG-----G-Q-----DVPSN-K-D--V-TSL----- 231  
Db 1888 EGVGSLYSVHEGIRGIPLD-P-NDKSDALVPVSGTSLAVGIDFHAENDTIYWDMDGLS 1945  
QY 232 -----D-W-----NS-----EG-TL--LATGS-Y--D-GF-----AR----- 250  
Db 1946 TISRAKDQWREDDVVVNGIGRVEGIADWIA-GNIYWTQGFVIEVARLNGSPRYVI 2004  
QY 251 -----I-----WTK-----DG-----N----- 257  
Db 2005 SOGLDKPRAITVHPEKGYLFWTEGHPRIERSRLDGTERRVVLNVNVSISWPNGISVDYOG 2064  
QY 258 --L--A-----STLG-----O-H-----K- 266  
Db 2065 GKLYWCDAEMDKIERIDLET-GENREVLSNNMDMFSVSVFEDFIYWSDRTHANGSIKR 2123  
QY 267 G-----P-----I--F----- 270  
Db 2124 GCKDNATDSVPLRTGIGVOLKDIKFENRDRQKXTNVCAVANGCQCLYRGGGORACAC 2183  
QY 271 A--L-----KW-----N-----KK----- 277  
Db 2184 AHGMLAEDGASCREYAGVLLYSERTILKSIHLSDERNLNAPVQPFEDPEHMKXVIALAFD 2243  
QY 278 --G-----N-F--I--LSAG--V--D--KTTII-----WDA----- 295  
Db 2244 YRAGTSPGTPNRIFFSDIHP--GNIQQINDGSGRTTIVENVGSVEGLAYHGRWDTLYWT 2301  
QY 296 -----HT-----G--E-----A--KO-Q--F----- 304  
Db 2302 SYTTITRHTVDQTRPGAFERETVITMSGDDHPRAFLVDELCEQNLMMFMNMLHPSIMR 2361  
QY 305 -----P-----F-----H-----SAPA 311  
Db 2362 AALSANVLTIEKDIRTPNGLAIDHRAEKLYFSDATLKDIERCEYDGSRRVILKSEVP 2421  
QY 312 ---LDV-----DW-----Q-----S 318



A;Cross-references: EMBL:U24703; NID:G902486; PID:G902487  
R.D'Arcangelo, G. EMBL Data Library, April 1995  
submitted to the EMBL Data Library, April 1995  
A;Reference number: S71844  
A;Accession: S71844  
A;Molecule type: mRNA  
A;Residues: 1-215, 'T', 217-1905, 'S', 1907-3355, 'V', 3357-3391, 'N', 3393-3461 <DA2>  
A;Cross-references: EMBL:U24703; NID:G902486; PID:G902487  
F:1-27/Domain: signal sequence #status predicted <SIG>  
F:28-3461/Product: reelin #status predicted <MAT>  
F:1769-1795/Domain: EGF homology <EGF>

Query Match 70.2%; Score 2568.8; DB 2; Length 3461;  
Best Local Similarity 11.8%; Pred. No. 1.5e-52;  
Matches 393; Conservative 92; Mismatches 16; Indels 2842; Gaps 361;

Qy	1	M-----SI-----S-----S-----D-----EV-----8
Db	1	MERGCWAPRALVLAVLLALLATLARAATGYVPFSPFFFLCTHGELEGDEQGEVLISL 60
Qy	9	N-----F-----LV-----YR-----LQ-----ES-----GF-----S-H 23
Db	61	HIAGNPTVYVPGQBYHVTISTSTFFDGLLVTLG-LYTSTSIQSSQIGSSAFAFGFMSDH 119
Qy	24	S-----S-----A-----F-----T-----F-----G-----I-----K-----31
Db	120	QFGNQFMCVSVASHVSHLPTNLSFVWIAPAGTCVNFATATHRQGVIFKDALAQOLC 179
Qy	32	SH-----I-----S-----Q-----SNI-----N-----GAL-----43
Db	180	EQGAPTEATAYSHLAEIHSDFVILRDDFDYSQVLELNPNIWVECSNCEMGQCGTMMGN 239
Qy	44	V-----P-----P-----A-----AL-----I-----S-----T-----I-----Q-----54
Db	240	AVTCEPYGPRELTTCLNTTASVLQFSIGSSCRFSYSDPSTVSYAKNNTADMIQLE 299
Qy	55	KG-----LQY-----V-----EA-----E-----VSN-----67
Db	300	KIRAPSNVSTVIHLYLPEEAKGESVQFMKQDSLVRGEVVEACWALDNLIV-INSahre 358
Qy	68	ED-----GT-----LP-----DG-----R-----76
Db	359	VLEDNLDPDVTGNWLFPGATVKHSCQSDGNSIYFHGNEGSFNFATTRDVLSTEDIQ 418
Qy	77	P-----P-----I-----ES-----LSLI-----DA-----86
Db	419	EQWSEEPESQPTGWDILGAVVGADCGTVESGLSLVFLKQGERKLCITPYMDTTGYGNLRFY 478
Qy	87	VM-----P-----D-----V-----V-----QT-----94
Db	479	FVMGICDPGVSHENDIILYAKIEGRKEHIALDTLTYSYKVPSPVSVVINPELQTPATK 538
Qy	95	Q-----RQ-----R-----D-----K-----LAQ-----Q-----106
Db	539	FCURQKSHQYNNRWAVDFHVLVPLPTSMHMIQFSINLGCCTHQPQNSVSLFSTNH 598
Qy	107	QA-----A-----AAA-----AA-----114
Db	599	GRSWLLHTECLPEICAGPHLPSTVYSSYSGWNRITIPNALTRDTRIRWQTGP 658
Qy	115	A-----A-----A-----A-----A-----A-----Q-----GS 123
Db	659	ILGNMWAIDNVYIGPCLKFCSGRGQCTRHCKCDPFGSPACEMASQTPPFMFISESPGS 718
Qy	124	AK-----N-----G-----E-----N-----129
Db	719	ARLSSVNFYSIRGAESVFCGVLASGKALVFNKGRRLQITSLDSSQSRFLQFTLRIG 778
Qy	130	T-----A-----NG-----E-----E-----N-----G-----137
Db	779	SKSVLSTCRAPDQPGEGVLLHYSDNGITWKLLEHYSYVNYHPRISIVELPDARQFGI 838
Qy	138	A-----A-----H-----T-----I-----AN-----N-----H-----145

Db	839	QFRWQPYTHSSQGEDVWAIDEIVMTSVLFNSISLDFTNLVEVTQSLGFLVGNVQPYCGHD 898
Qy	146	T-----D-----M-----MEVDG-----D-----V-----E-----156
Db	899	WTLCTGCDKSLASSMYVETQSMQI-GASYMIQPSLVMGCGQKYTPHMDNQVKLEYSANTH 957
Qy	157	I-----I-----P-----PN-----K-----A-----VVL-----R-----166
Db	958	GLTWHLVQBECLPMPSQBEFTSASIYHASEFTQMRVRVTVPVLPQKTSQATRFRWSQSY 1017
Qy	167	CH-----GH-----CH-----E-----SE-----V-----F-----173
Db	1018	TAQDEWALDNIIYGOCPNMCSCGHGSCDHGVCRCDOGYGTECHPEALPSTIMSDFENP 1077
Qy	174	I-----I-----C-----A-----W-----177
Db	1078	SSWESDQVEVIGVEVVKPEQCGVSSGSLYFSKAGKQLVSWDLDTSMVDVQFVQI 1137
Qy	178	N-----N-----P-----V-----SD-----L-----L-----ASG-----187
Db	1138	GGESAACNKPDSREEGILLQYSNNGIOHLLAEMFYSDFSKPRFVYLEPRAAGTPTCTR 1197
Qy	188	SG-----D-----S-----T-----A-----R-----IW-----196
Db	1198	FRWKVPVFGEDYDQWAVDDIILSEKQKQVTPVNPVTLQPQFYKPADPYPMNQMSWL 1257
Qy	197	N-----N-----LSEN-----S-----T-----S-----G-----S-----Q-----L-----209
Db	1258	MLANEGMAKNDSFCATTPSAMVFGSDGRFAVTRDLTLKPGVVLQFKLNIQCTQSFSST 1317
Qy	210	V-----LR-----H-----C-----IR-----EG-----GOD-----V-----222
Db	1318	APVLLQYSHDAGMFWLLKEGCFPASAKGCEGNSRELSEPTVYTYG-DFEETRITIAI 1376
Qy	223	P-----S-----N-----K-----D-----227
Db	1377	PRSLASKTRFRWIOBSSQKNVPPGLDGVVISBPSPSCYSGHGDICISGVCFDLGYTA 1436
Qy	228	V-----V-----T-----S-----LD-----W-----NSBG-----TL-----L-----A-----241
Db	1437	AGTCTVSNTPNHSEMPDRPEGLKPLWYKITGGVGTGCTGLN-DGRSLYFNLGLKREAR 1495
Qy	242	T-----GS-----Y-----D-----G-----F-----248
Db	1496	TVPLDTRNLSLVQFYIQGSKTSGITYITPRARYEGLVVOYNDNGILMHLRLDLMFMSF 1555
Qy	249	A-----RiW-----T-----K-----DG-----256
Db	1556	LEPQIISIDLPREAKTPATAFR-WWQPOGHKHAQWALGDVLIGVNDSSQTGFQDKLDGS 1614
Qy	257	N-----LAS-----T-----L-----G-----Q-----H-----KG-----267
Db	1615	IDLQANWRIQGVDDIDCL-SMDTALIFTENIGNPRYAETWDFHVSESSFLQWEMNMG 1673
Qy	268	PIF-----A-----L-----K-----W-----NKK-----G-----278
Db	1674	SKP-FSGAHGIGLOQLSLNNGKDWLVTECVPTTIGCVHYTESYTSRPFQWRRTVY 1732
Qy	279	N-----P-----I-----LSAGVDKT-----TI-----I-----W-----DA-----295
Db	1733	LPLATNSPRTRFRWITQNTYTVGAD-SWAIDNVILASGCPMCSGRGICDSGRVCVDRGPG 1791
Qy	296	H-----H-----T-----GE-----299
Db	1792	GFPCVVPVPLPSILKDDFNGLHPLDMPVEVYGAERGNLNGETIKSGTCLIFKEGLRMLI 1851
Qy	300	AK-----Q-----QF-----P-----F-----H-----SA-----309
Db	1852	SRDLCTNTMYVQFSRLFIKATPTPERSHILLQFVSQGVTHLMDPEYFPQTTRILFIN 1911
Qy	310	P-----A-----L-----D-----V-----D-----W-----H-----316



Db 1912 VPLPYGAQTNRFLRWQPYNNGKKEEIIWIDDIIIDGNLNNPVLILLDTDFGPREDNW 1971  
Qy 317 -----Q--S-N-NT----- 321  
Db 1972 FFYPGGNIGLYCPSYSGKAGEEDSAMVFSVNEGEHSIITRDLNVNENTIIQEFINVCS 2031  
Qy 322 -----F-A-----S-CST----- 327  
Db 2032 TDSSADPVRLEPSRDFGATWHLPLCYHSSLSVSLCSTEHPSSTYYAGTTOGWRE 2091  
Qy 328 -----D-----MC-IH-----VC 334  
Db 2092 VHFGLHLGCVSRFRWYQYFAGSQPVTAIDNVYIGQCEMICYGHGSCINGTKCIC 2151  
Qy 335 -----K--L--GO--DR-----P-----I-KT-----F-Q-G- 347  
Db 2152 DPGYSGPTCKISTKNDPDLKDDFEGGLESDRFLMSGGRPSKCGILSSGNLFFNEDGL 2211  
Qy 348 -----H-----T-N-----EV----- 352  
Db 2212 RMLVTRDLDLSHARFVQFMRLGCGKVPDRPSQPVLLQYSLNGLSLWSLLQEFLEFSNS 2271  
Qy 353 NA--I-----K-----W----- 357  
Db 2272 NVGRYIALEMPKARSGLRWRWQPSENGHFYSPWVIDQILIGNIGNTVLEDDFSTL 2331  
Qy 358 D-----P--TG-----N--LL-----ASCS--D 369  
Db 2332 DSRKLLHPGCTKMPVCGTGDALVIEKASTRYVTTDIANNEDSFLOIDFAASCSTVD 2391  
Qy 370 -----DM-----T-----L-KI-----W----- 376  
Db 2392 SCVAIELEYSVDLGLSWHPLVRDCLPTNVECSRHYLQRLIVSDTFNKWTRITLPLPSYTR 2451  
Qy 377 S-----M-----KQ-----DN-----CV-----H-----D-- 386  
Db 2452 SQATFRWHQAPFPDKQOTWAIIDNVYIGDCLDMCSGHRVCQVQSCVDEQWGLYCDER 2511  
Qy 387 -----L--Q-----O-----H----- 390  
Db 2512 ETSLPTQLKDNFRAPSNQNLTVSGKLSVCGKLVASGLALHFGSGCSRLLVTVDLNLIT 2571  
Qy 391 NKE-I-----Y-----T-----IKW-----S-PTG-----P 404  
Db 2572 NAEFIQYFYMGCLITPSNRNQGLLEYSVNGGITWLLMEIFYDQYKXP-GFVNILLPP 2630  
Qy 405 -----GT-----NNP-----N--A--NLMLA-SA-----SF-----D 421  
Db 2631 DAKBIATFRFWMQ--PRHDGLDQNDWAIIDNVLISSADQRTVMLDTFSSAPVQHERSPAD 2689  
Qy 422 S-TV-R-----L-----W--DVD--R-----GI--C-IH--TL--TKH- 441  
Db 2690 AGEVGRIAEMFLEDKTSVNNWLFHD-DCTVERFCSDPDGVNLCGSHDGREVIYAVT-HD 2747  
Qy 442 -----Q-----E-P-----V-YSAF-----S-----P-D----- 453  
Db 2748 LPTENWIMQKISVCKVPEKIAQNIHQVSTDFGVSWSYLVPOCLPADPKCSGSVSQ 2807  
Qy 454 -----G--R--Y-L-AS-G-----S-----F-----D- 463  
Db 2808 PSVFFTEGWKRTITYPLPESLITGNVPRFRFYQYSDVQWAIIDNFYLGPGCLDNCGGHDC 2867  
Qy 464 -K--CV-----H-----I-----W-----NT- 471  
Db 2868 LKEQCICDPGSGPNCYLTHSLKTLKERFDSSEIKPDLMMLEGGSTCTCEGVLAENTA 2927  
Qy 472 -----Q--T-----GA--L-----V-----H-----S----- 479  
Db 2928 LYFGSTVRAITQDLDLRGAKFLOVWGRIGSENNTSCHRVPCKRKEGVLLDFSTGGIT 2987  
Qy 480 -----Y--R-----G-----T-----G-----GI----- 486  
Db 2988 WTLLEHMDQKYSVRHVDVILLPEGALTNTTLRWMPFVINGNLVSVGVGACAVGTGQH 3047

Qy 487 -----F-----E-----V-----C-----W-----NA 493  
Db 3048 SDWWSRHHPSQLVDTFDDSGSSHEENWSFYPNVARTAGCGNPSFHYLWPNKKKDKTHNA 3107  
Qy 494 -----A-----GD-----K----- 497  
Db 3108 LSSRELIIOGYMMQFKIIVVGCEATSCGDLHLSVMLEYTKDARSDSMQLVQTQCLPSSNS 3167  
Qy 498 VG-----A-----SA-----S--D-- 504  
Db 3168 IGCSPFPQFHEATIYNVNSSWKRTITQLPDHVSSATQFRWIOKGEETEKSQSWAIDHVV 3227  
Qy 505 -G-----S-----VC-----LDL-R--K 514  
Db 3228 IGEACPKLCSGHGYCTTGAVCICDESFQGDGCSVFSHELPSYIKDNFESARVTEANWETI 3287  
Qy 509 -----V-----LDL-R--K 514  
Db 3288 QGCVIGSGCGQLAPYAHGDSLYFNGCQIRQAATKPLDLTRASK 3330

Search completed: January 3, 2005, 15:26:59  
Job time : 101.667 secs

**This Page Blank (uspto)**



RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,  
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,  
RA Togashi T., Oyama M., Hata H., Watanabe M., Komatsu I.,  
RA Mizushima-Sugano J., Sato T., Shirai Y., Takahashi Y., Nakagawa K.,  
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maehuo Y., Yamashita R.,  
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
RT "Complete sequencing and characterization of 21,243 full-length human  
RT cDNAs.";  
RL Nat. Genet. 36:40-45 (2004).  
RL [4]  
RN SEQUENCE OF 1-125 FROM N.A.  
RP TISSUE=Adrenal cortex;  
RX MEDLINE=22398257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Vallalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
RN [5]  
RP COMPONENT OF THE N-COR COMPLEX WITH TBL1X; CORO2A AND HDAC3, AND  
RP HISTONE-BINDING.  
RX PubMed=12628926;  
RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,  
RA Wong J.;  
RT "Purification and functional characterization of the human N-Cor  
RT complex: the roles of HDAC3, TBL1 and TBL1X.";  
RL EMBO J. 22:1336-1346 (2003).  
RN [6]  
RP FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.  
RX PubMed=14980219;  
RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;  
RT "A corepressor/coactivator exchange complex required for  
RT transcriptional activation by nuclear receptors and other regulated  
RT transcription factors.";  
RL Cell 116:511-526 (2004).  
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the  
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated  
CC transcription units. Plays an essential role in transcription  
CC activation mediated by nuclear receptors. Probably acts as  
CC integral component of the N-Cor corepressor complex that mediates  
CC the recruitment of the 19S proteasome complex, leading to the  
CC subsequent proteasomal degradation of N-Cor complex, thereby  
CC allowing cofactor exchange, and transcription activation.  
CC -!- SUBUNIT: Component of the N-Cor repressor complex, at least  
CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.  
CC Probable component of some E3 ubiquitin ligase complex. Interacts  
CC with histones H2B and H4.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- TISSUE SPECIFICITY: Ubiquitous.  
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and  
CC apparently displays the same function as component of ubiquitin E3  
CC ligase complexes (By similarity).  
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.  
CC -!- SIMILARITY: Contains 1 F-box-like domain.  
CC -!- SIMILARITY: Contains 1 Lish domain.  
CC -!- SIMILARITY: Contains 8 WD repeats.  
CC -!- CAUTION: Ref.4 sequence differs from that shown due to a potential  
CC poly-A sequence starting in position 125.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
CC EMBL; AF314544; AAK00301.1; --  
CC EMBL; AF268193; AAG44736.1; --  
CC EMBL; AK022956; BAB14331.1; --  
CC EMBL; BC060320; BAH60320.1; ALT\_TERM.  
CC HSSP; P16649; 1ERJ.  
CC MIM; 608628; --  
CC InterPro; IPR006594; Lish.  
CC InterPro; IPR011047; Quin\_alc\_DH\_like.  
CC InterPro; IPR001680; WD40.  
CC Pfam; PF04000; WD40; 8.  
CC PRINTS; PR00320; GPROTEINERPT.  
CC ProDom; PD000018; WD40; 3.  
CC SMART; SM00667; Lish; 1.  
CC SMART; SM00320; WD40; 8.  
CC PROSITE; PS00896; Lish; 1.  
CC PROSITE; PS00678; WD\_REPEATS\_1; 4.  
CC PROSITE; PS00892; WD\_REPEATS\_2; 6.  
CC PROSITE; PS0294; WD\_REPEATS\_REGION; 1.  
CC Activator; Nuclear protein; Repeat; Transcription regulation;  
CC Ub1 conjugation pathway; WD repeat.  
CC FT DOMAIN 4 36  
CC FT DOMAIN 41 86 F-box-like.  
CC REPEAT 167 206 WD 1.  
CC REPEAT 223 262 WD 2.  
CC REPEAT 264 303 WD 3.  
CC REPEAT 306 344 WD 4.  
CC REPEAT 347 386 WD 5.  
CC REPEAT 389 437 WD 6.  
CC REPEAT 440 479 WD 7.  
CC REPEAT 481 513 WD 8.  
CC FT DOMAIN 108 118 Poly-Ala.  
CC FT CONFLICT 31 31 E -> K (in Ref. 2).  
CC FT CONFLICT 59 59 Y -> H (in Ref. 3).  
CC FT CONFLICT 389 389 A -> Q (in Ref. 2).  
CC SQ SEQUENCE 514 AA; 55595 MW; 0B556D2EE4BA796D CRC64;  
CC  
CC Query Match 99.8%; Score 3650; DB 1; Length 514;  
CC Best Local Similarity 99.8%; Pred. No. 1.6e-65;  
CC Matches 512; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
CC  
CC QY 1 MSISDEVNPLVRYLQESGFHSFTFGIKSHISQSNINGALVPPAALISIIQGLQYV 60  
CC DB 1 MSISDEVNPLVRYLQESGFHSFTFGIESHSQSNINGALVPPAALISIIQGLQYV 60  
CC  
CC QY 61 EAEVSINEDGTLDPGRPIESLSLIDAMPDVVTRQAVRDVKLAQQQAAAAAASQ 120  
CC DB 61 EAEVSINEDGTLDPGRPIESLSLIDAMPDVVTRQAVRDVKLAQQQAAAAAASQ 120  
CC  
CC QY 121 QGSAKNGENTANGENGATIANNTDMMEDVDGVEIPPNKAVLVGRHSEVEFICAWN 180  
CC DB 121 QGSAKNGENTANGENGATIANNTDMMEDVDGVEIPPNKAVLVGRHSEVEFICAWN 180  
CC  
CC QY 181 SLLASGSGDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWSNSETLL 240  
CC DB 181 SLLASGSGDSTARINWLSNSTSGSTQLVLRHCIREGGQDVPSNKDVTSLDWSNSETLL 240  
CC  
CC QY 241 ATGSYDGFARIWTKDGNLSTLQGHKGPFPALKWNKGNFILSAGVDKTTIIINDAHTGEA 300  
CC DB 241 ATGSYDGFARIWTKDGNLSTLQGHKGPFPALKWNKGNFILSAGVDKTTIIINDAHTGEA 300  
CC  
CC QY 301 KQOFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360  
CC DB 301 KQOFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360

QY 361 GNLLASCSDDMTLKWSMKQDNCVHDLQHQHKEIYTIKWSPTGPTNNPNANMLASASF 420  
DB 361 GNLLASCSDDMTLKWSMKQDNCVHDLQHQHKEIYTIKWSPTGPTNNPNANMLASASF 420  
QY 421 DSTVRLWDVDRGICHTLTKHQPVSVAESPGRVYSLASGSPKCVHINTQTGALVHSY 480  
DB 421 DSTVRLWDVDRGICHTLTKHQPVSVAESPGRVYSLASGSPKCVHINTQTGALVHSY 480  
QY 481 RTGGTGFIFVCVNAAGDKVGSASDGSVCVLDLRK 514  
DB 481 RTGGTGFIFVCVNAAGDKVGSASDGSVCVLDLRK 514

RESULT 2

TBLR MOUSE STANDARD: PRT; 514 AA.

AC Q8BHJ5; Q8CEG4; Q8VEG3; Q9EQD4;  
DT 05-JUL-2004 (Rel. 44, Created)  
DT 05-JUL-2004 (Rel. 44, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE F-box-like/WD-repeat protein TBLR1 (Nuclear receptor corepressor/HDAC3  
DE complex subunit TBLR1) (TBL1-related protein 1).  
GN Name=Tblr1; Synonyms=Iral;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c;  
RA Zhang X., Dormady S., Basch R.;  
RT "Identification of four human cDNAs that are differentially expressed  
RT by early hematopoietic progenitors."  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Testis;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojibori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanpin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt D., Brusci V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmer S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashina T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sadelain A., Schneider C., Semple C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,  
RA Verardo R., Wagner L., Wahlstedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wynshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sakaki D., Shibata K., Shingawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).  
RN [3]  
RP SEQUENCE OF 314-514 FROM N.A.  
RC TISSUE=Breast tumor;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,

RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Uesdin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smalish D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences."  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the  
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated  
CC transcription units. Plays an essential role in transcription  
CC activation mediated by nuclear receptors. Probably acts as  
CC integral component of the N-CoR corepressor complex that mediates  
CC the recruitment of the 19S proteasome complex, leading to the  
CC subsequent proteosomal degradation of N-CoR complex, thereby  
CC allowing cofactor exchange, and transcription activation (By  
CC similarity).  
CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least  
CC composed of NCOR1, NCOR2, HDAC3, TBLR1, TBLR2, and GPS2.  
CC Probable component of some E3 ubiquitin ligase complex. Interacts  
CC with histones H2B and H4 (By similarity).  
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and  
CC apparently displays the same function as component of ubiquitin E3  
CC ligase complexes (By similarity).  
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.  
CC -!- SIMILARITY: Contains 1 F-box-like domain.  
CC -!- SIMILARITY: Contains 8 WD repeats.  
CC -!- SIMILARITY: Contains 8 WD repeats.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL: AF268195; AAG44738.1; -  
CC EMBL: AK029595; BAC26526.1; -  
CC EMBL: AK033347; BAC28241.1; -  
CC EMBL: AK036064; BAC29294.1; -  
CC EMBL: BC018512; AAH18512.1; -  
CC HSSP: P16649; 1ERJ.  
CC InterPro: IPR006594; Lish.  
CC InterPro: IPR011047; Quin\_alc\_DH\_like.  
CC InterPro: IPR001680; WD40.  
CC Pfam: PF00400; WD40; 8.  
CC PRINTS: PR00320; GPROTEINBRPT.  
CC ProDom: PD000018; WD40; 3.  
CC SMART: SM00667; Lish; 1.  
CC SMART: SM00320; WD40; 8.  
CC PROSITE: PS50896; Lish; 1.  
CC PROSITE: PS00678; WD\_REPEATS\_1; 4.  
CC PROSITE: PS50082; WD\_REPEATS\_2; 6.  
CC PROSITE: PS50294; WD\_REPEATS\_REGION; 1.  
CC Activator; Nuclear protein; Repeat; Transcription regulation;  
KW Ub1 conjugation pathway; WD repeat.  
FT DOMAIN 4 36 Lish.  
FT DOMAIN 41 86 F-box-like.  
FT REPEAT 167 206 WD 1.  
FT REPEAT 223 262 WD 2.  
FT REPEAT 264 303 WD 3.

```

FT REPEAT      306 344      WD 4.
FT REPEAT      347 386      WD 5.
FT REPEAT      389 437      WD 6.
FT REPEAT      440 479      WD 7.
FT REPEAT      481 513      WD 8.
FT DOMAIN      108 117      Poly-Ala.
FT CONFLICT    185 185
FT CONFLICT    374 374      A -> V (in Ref. 1).
FT CONFLICT    480 480      X -> R (in Ref. 2; BAC29294).
FT CONFLICT    480 480      Y -> C (in Ref. 3).
SQ SEQUENCE    514 AA; 55661 MW; 13BEC1C2C7FBFB14 CRC64;

Query Match      98.9%; Score 3618; DB 1; Length 514;
Best Local Similarity 98.6%; Pred. No. 7.5e-65;
Matches 507; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 MSISDEVNVLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNVLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQKGLQYV 60

Qy 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQVTRQOAYRDKLAQQQAAAAAASQ 120
Db 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQVTRQOAYRDKLAQQQAAAAAATNQ 120

Qy 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFIKAWNPV 180
Db 121 QGSAKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEVFIKAWNPV 180

Qy 181 SLLASGSGDSTARIMNLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLDWNSGTLT 240
Db 181 SLLASGSGDSTARIMNLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLDWNSGTLT 240

Qy 241 ATGSYDGFARITWKDGNLSTLQHGKGPFIKALKNKNGFILSAGVDKTTIINDAHTGEA 300
Db 241 ATGSYDGFARITWKDGNLSTLQHGKGPFIKALKNKNGFILSAGVDKTTIINDAHTGEA 300

Qy 301 KQFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 360
Db 301 KQFPFHSAPALVDWQSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEVNAIKWDPT 360

Qy 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYIKWSPGPTGNTNPNANMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQHNKEIYIKWSPGPTGNTNPNANMLASAF 420

Qy 421 DSTVRLWDVDRGICHTLTKHQPVSVAFSPDGRVYLAGSFKCVHIWNTQTGALVHSY 480
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFSPDGRVYLAGSFKCVHIWNTQTGALVHSY 480

Qy 481 RTGGIPEVCWNAAGDKVGSASDGSVCVLDLRK 514
Db 481 RTGGIPEVCWNAAGDKVGSASDGSVCVLDLRK 514

```

## RESULT 3

```

Q6GPC6 ID Q6GPC6 PRELIMINARY; PRT; 522 AA.
AC Q6GPC6;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Hypothetical protein.
OC Xenopus laevis (African clawed frog).
OC Amphibia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,

```

```

RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Donald M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McSwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzyvinski M.I., Skalek U., Smailus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
RT initiative.";
RL Dev. Dyn. 225:384-391 (2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.;
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: Contains 8 WD repeats.
DR EMBL: BC073215; AAH73215.1; -.
DR InterPro: IPR006594; Lish.
DR InterPro: IPR011047; Quin_alc_DH_like.
DR InterPro: IPR001680; WD40.
DR Pfam: PF00400; WD40; 8.
DR PRINTS: PR00320; GPROTEINBRPT.
DR ProDom: PD000018; WD40; 3.
DR SMART: SM00667; Lish; 1.
DR SMART: SM00320; WD40; 8.
DR PROSITE: PS00896; LISH; 1.
DR PROSITE: PS00678; WD_REPEATS_2; 6.
DR PROSITE: PS00824; WD_REPEATS_1; 4.
DR PROSITE: PS00294; WD_REPEATS_REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 522 AA; 56284 MW; 4FB0F58F03B461F CRC64;

```

```

Query Match      98.1%; Score 3589.2; DB 2; Length 522;
Best Local Similarity 96.4%; Pred. No. 3.2e-64;
Matches 503; Conservative 5; Mismatches 6; Indels 8; Gaps 3;

```

```

Qy 1 MSISDEVNVLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISDEVNVLVRYLQESGFSHSAFTGKISHISQSNINGALVPPAALISIIQKGLQYV 60

Qy 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQVTRQOAYRDKLAQQQ-----AAAAAAA 116
Db 61 EAEVSNEDGTLFDGRPIESLSLIDAVMPDVQVTRQOAYRDKLAQQQAAAAAAA 120

Qy 117 AA-S---QGSARKNGENTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEV 172
Db 121 AATSANNQPPAKNGESTANGENGAAHTIANNHTDMVEVDGVEIPPNKAVVLRGHESEV 180

Qy 173 FICANNPVSDLLASGSGDSTARIMNLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLD 232
Db 181 FICANNPVSDLLASGSGDSTARIMNLSNSTSGTQVLVLRHCIRREGQDVPSNKDVTSLD 240

Qy 233 WNSEGTLLATGSDGFARITWKDGNLSTLQHGKGPFIKALKNKNGFILSAGVDKTTI 292
Db 241 WNSEGTLLATGSDGFARITWKDGNLSTLQHGKGPFIKALKNKNGFILSAGVDKTTI 300

Qy 293 WDAHTGEAKQPPHSPALVDWQSNNTFASCSTDMCIHVCKLQGDRIPIKTFQGHTEV 352

```

```
Db 301 WDAHTGEAKQPPHSPALPDVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEV 360
Qy 353 NAIKWDPTGNLLASCSDMTLTKIWSMKQDNCVHDLQOHNEIYTIKWSPTGPGTNNPNAN 412
Db 361 NAIKWDPTGNLLASCSDMTLTKIWSMKQDNCVHDLQAHNEIYTIKWSPTGPGTNNPNAN 420
Qy 413 LMLASAFDSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYLASGSPDKCVHIWNTQ 472
Db 421 LMLASAFDSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYLASGSPDKCVHIWNTQ 480
Qy 473 TGAHVHSGTGGIFEVCAAGDKVGSASDGSVCVLDLRK 514
Db 481 TGAHVHSGTGGIFEVCAAGDKVGSASDGSVCVLDLRK 522

RESULT 4
Q7SZM9 PRELIMINARY; PRT; 519 AA.
ID Q7SZM9 ID Q7SZM9 STANDARD; PRT; 527 AA.
AC Q7SZM9 AC Q7SZM9 Q8BMM0; Q8BYQ4; Q8COA1;
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Nuclear receptor co-repressor complex subunit TBLR1.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RA Tomita A., Buchholz D.R., Obata K., Shi Y.B.,
RA "Fusion Protein of Retinoic Acid Receptor (alpha) with Promyelocytic
RT Leukemia Protein or Promyelocytic Leukemia Zinc Finger Protein
RT Recruits N-CoR-TBLR1 Corepressor Complex to Repress Transcription in
RT Vivo.";
RL J. Biol. Chem. 278:30788-30795 (2003).
CC -1- SIMILARITY: Contains 8 WD repeats.
DR EMBL; AY225088; AAP20646.1; -.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR006594; L1SH.
DR Pfam; PF00400; WD40; 8.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; L1SH; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; L1SH; 1.
DR PROSITE; PS00678; WD REPEATS_1; 4.
DR PROSITE; PS00082; WD REPEATS_2; 6.
DR PROSITE; PS00294; WD REPEATS_REGION; 1.
KW Receptor; Repeat; WD repeat.
SQ SEQUENCE 519 AA; 56043 MW; 5E998EDC8C92296 CRC64;

Query Match 97.9%; Score 3580.5; DB 2; Length 519;
Best Local Similarity 96.5%; Pred. No. 4.8e-64;
Matches 501; Conservative 7; Mismatches 6; Indels 5; Gaps 2;

Qy 1 MSISSEVNVFLVRYLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSISSEVNVFLVRYLOESGFSHSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQYV 60
Qy 61 EAEVSVNEDGTLFDGRIEISLIDAVPDPVQTRQAVYRDKLAQQQ--AAAAA 118
Db 61 EAEVSVNEDGTLFDGRIEISLIDAVPDPVQTRQAVYRDKLAQQQATAAAAAA 120
Qy 119 S---QQGSAKNGENTANGEAGHTIANNHTDMMEYDGDVIEIPPKAVLRLGHSEVFIC 175
Db 121 TPNNQPPAKNGENTANGEAGHTIANNHTDMMEYDGDVIEIPSKAVLRLGHSEVFIC 180
Qy 176 ANPVSDDLASGSDSTARIWNLSENSTSGSTQVLRLHCTIREGGQDVPSNKDVTSLDWNS 235
Db 181 ANPVSDDLASGSDSTARIWNLSENSTSGSTQVLRLHCTIREGGQDVPSNKDVTSLDWNS 240
```

```
Qy 236 EGTLLATGSDGFARITWTKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTIIWDA 295
Db 241 EGTLLATGSDGFARITWTKDGNLSTLQHGKPIFALKWKNKGNFILSAGVDKTTIIWDA 300
Qy 296 HTGEAKQPPHSPALPDVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAI 355
Db 301 HTGEAKQPPHSPALPDVDWQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAI 360
Qy 356 KWDPTGNLLASCSDMTLTKIWSMKQDNCVHDLQOHNEIYTIKWSPTGPGTNNPNANLML 415
Db 361 KWDPTGNLLASCSDMTLTKIWSMKQDNCVHDLQAHNEIYTIKWSPTGPGTNNPNANLML 420
Qy 416 ASAFDSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYLASGSPDKCVHIWNTQTA 475
Db 421 ASAFDSTVRLWDVDRGICHTLTKHQEPVYVAFSPDGRYLASGSPDKCVHIWNTQTA 480
Qy 476 LVHSGTGGIFEVCAAGDKVGSASDGSVCVLDLRK 514
Db 481 LVHSGTGGIFEVCAAGDKVGSASDGSVCVLDLRK 519

RESULT 5
TBLX_MOUSE STANDARD; PRT; 527 AA.
ID TBLX_MOUSE ID TBLX_MOUSE STANDARD; PRT; 527 AA.
AC Q9QX87; Q8BMM0; Q8BYQ4; Q8COA1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein TBLX (Transducin beta-like 1X protein).
GN Name=Tblx; Synonyms=Tb11;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J;
RA TISSUE=Hypothalamus, Medulla oblongata, and Pituitary;
RA MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Koenig A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Muralis L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenawa Y., Taylor M.S., Teasdale R.D., Tomita K.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wyszynski-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Haehizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shingawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6; TISSUE=Brain;
RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
```

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diczienko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stawton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Lequellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fanev J., Helton E., Kettman M., Madan A.C., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A.C., Young J.W., Green E.D., Dickson M.C.,  
 RA Blakesley R.W., Touchman J.W., Schmutz J., Myers R.M.,  
 RA Butterfield V.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RA "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE OF 200-527 FROM N.A.  
 RC STRAIN-129/SvJ;  
 RA Botcherby M.R.M., Straw R., Clarke D., Greystrom J.S., Weston P.,  
 RA Hunter G., Kimberley C., Rhodes M.;  
 RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP TISSUE SPECIFICITY.  
 RX MEDLINE=92264241; PubMed=10330347;  
 RA Bassi M.T., Ramasar R.S., Caciotti B., Winship I.M., De Grandi A.,  
 RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.;  
 RT "X-linked late-onset sensorineural deafness caused by a deletion  
 RL involving OAI and a novel gene containing WD-40 repeats.";  
 CC Am. J. Hum. Genet. 64:1604-1616(1999).  
 CC -!- FUNCTION: F-box-like protein involved in the recruitment of the  
 CC ubiquitin/19S proteasome complex to nuclear receptor-regulated  
 CC transcription units. Plays an essential role in transcription  
 CC activation mediated by nuclear receptors. Probably acts as  
 CC integral component of corepressor complexes that mediates the  
 CC recruitment of the 19S proteasome complex, leading to the  
 CC subsequent proteosomal degradation of transcription repressor  
 CC complexes, thereby allowing cofactor exchange (By similarity).  
 CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least  
 CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.  
 CC Component of an E3 ubiquitin ligase complex containing UBE2D1,  
 CC STAH1, CACYPB/STP, SKP1A, APC and TBL1X. Probably part of other  
 CC corepressor complexes, that do not contain NCOR1 and NCOR2.  
 CC Interacts with histones H2B, H3a and H4 (By similarity).  
 CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -!- TISSUE SPECIFICITY: Expressed in the cochlea.  
 CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and  
 CC apparently displays the same function as component of ubiquitin E3  
 CC ligase complexes (By similarity).  
 CC -!- SIMILARITY: Belongs to the WD-repeat E3 family.  
 CC -!- SIMILARITY: Contains 1 F-box-like domain.  
 CC -!- SIMILARITY: Contains 8 WD repeats.  
 CC -!- SIMILARITY: Contains 8 WD repeats.  
 CC -!- CAUTION: Ref.3 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AK030547; BAC27015.1; -  
 CC DR EMBL; AK031937; BAC27612.1; -  
 CC DR EMBL; AK038674; BAC30092.1; -  
 CC DR EMBL; BC043105; AAH43105.1; -  
 CC DR EMBL; F38006; CAB61534.1; ALT\_SEQ.  
 CC MGD; MGI:1336172; Tbl1x.

DR InterPro: IPR001690; WD40.  
 DR Pfam: PF004400; WD40; 5.  
 DR PRINTS: PR00320; GPROTEINRPT.  
 DR PRODOM: PD000018; WD40; 2.  
 DR SMART: SM00320; WD40; 5.  
 DR PROSITE: PS00896; LISH; 1.  
 DR PROSITE: PS00678; WD\_REPEATS\_1; 4.  
 DR PROSITE: PS00082; WD\_REPEATS\_2; 6.  
 DR PROSITE: PS0294; WD\_REPEATS\_REGION; 1.  
 KW Activator; Nuclear protein; Repeat; Transcription regulation;  
 KW Ub1 conjugation pathway; WD repeat.  
 FT DOMAIN 4 36 Lish.  
 FT DOMAIN 41 86 F-box-like.  
 FT REPEAT 180 219 WD 1.  
 FT REPEAT 236 275 WD 2.  
 FT REPEAT 277 316 WD 3.  
 FT REPEAT 319 359 WD 4.  
 FT REPEAT 360 399 WD 5.  
 FT REPEAT 402 450 WD 6.  
 FT REPEAT 453 492 WD 7.  
 FT REPEAT 494 526 WD 8.  
 FT DOMAIN 108 121 Poly-Ala.  
 FT DOMAIN 130 133 Poly-Ala.  
 FT CONFLICT 104 104 A -> T (in Ref. 1; BAC27015).  
 FT CONFLICT 349 349 L -> H (in Ref. 1; BAC27612).  
 SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFDA6 CRC64;  
 Query Match 93.4%; Score 3418.1; DB 1; Length 527;  
 Best Local Similarity 88.1%; Pred. No. 1.4e-60;  
 Matches 467; Conservative 29; Mismatches 15; Indels 19; Gaps 8;  
 QY 1 MSISDEVNPLVYRYLOESGFSHSAFTGKISHISOSNINGALVPPAALISILQKGLQYV 60  
 DB 1 MSISDEVNPLVYRYLOESGFSHSAFTGKISHISOSNINGALVPPAALISILQKGLQYV 60  
 QY 61 EAEVSNEDGTLFDGRIEISLSLIDAVMDVQVTRQQAAYRDKLAQQA-AAAAAAAA- 118  
 DB 61 EAEISINEDGTVDGRIEISLSLIDAVMDVQVTRQQAAYRDKLAQQA-AAAAAAAA- 120  
 QY 119 --S-----QQSAXNGENTANGENGAAHTANNNTDM-MEVDGVEIPPNKAVL 165  
 DB 121 ATSTAATTPAAAQQNPKNKGATVNGEENGAAHAI--NNHS-KPMEIDGSEIIPPSKATVL 178  
 QY 166 RGHESEVFICANNPVSDDLASGSGSTARIWNLSENSTSGTQLVLRHCIRREGQDVPSN 225  
 DB 179 RGHESEVFICANNPVSDDLASGSGSTARIWNLSENSTSGTQLVLRHCIRREGQDVPSN 238  
 QY 226 KDVTSLDWNSEGLLATGSDYDGFARITWKDGNLSTLQGHKGPFPALKWNKGNFTLSAG 285  
 DB 239 KDVTSLDWNSEGLLATGSDYDGFARITWKDGNLSTLQGHKGPFPALKWNKGNFTLSAG 298  
 QY 286 VDKTTIINDAHTGEAKQOPFFHSAPALVDVQSN--TFASCSTDMCIHVCKLQDRPIKT 344  
 DB 299 VDKTTIINDAHTGEAKQOPFFHSAPALVDVQ--NNTTFASCSTDMCIHVCKLQDRPVKT 357  
 QY 345 FQGHTEVNAIKWDPTGNLLASCDSDMTLKIWSMKQDNCHDLQOHKKEIYTIKWSPTGP 404  
 DB 358 FQGHTEVNAIKWDPTGNLLASCDSDMTLKIWSMKQDNCHDLQOHKKEIYTIKWSPTGP 417  
 QY 405 GTNNPNANIMLASGFSDFSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGFDK 464  
 DB 418 ATSNPNINMLASGFSDFSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGFDK 477  
 QY 465 CVHIWNTQTGALVSHVSGTGGIFEVCMNADGKVGASASDGSVCVLDLRK 514  
 DB 478 CVHIWNTGSLVSHVSGTGGIFEVCMNARGDKVGASASDGSVCVLDLRK 527  
 RESULT 6  
 TBLX HUMAN  
 ID TBLX HUMAN STANDARD; PRT; 526 AA.  
 AC O60907; Q86UY2;  
 DT 16-OCT-2001 (Rel. 40, Created)



DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DE 05-JUL-2004 (Rel. 44, Last annotation update)  
DE F-box-like/WD-repeat protein TBL1X (Transducin beta-like 1X protein)  
DE (Transducin-beta-like 1, X-linked) (SWAP55).  
CN Name=TBL1X; Synonyms=TBL1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RN SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DISEASE.  
RX MEDLINE=99264241; PubMed=10330347;  
RA Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,  
RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.,  
RT "X-linked late-onset sensorineural deafness caused by a deletion  
RT involving OAI and a novel gene containing WD-40 repeats.";  
RL Am. J. Hum. Genet. 64:1604-1616(1999).  
RN [2]  
RN SEQUENCE FROM N.A.  
RC TISSUE=Lymph, and Pancreas;  
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
RA Strausberg R.L., Peingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
RA Altschul S.F., Zebberg B., Buetow K.H., Schaefer C.P., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Diatchenko M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ustin T.B., Toshiyuki S., Carninci P., Prange C.,  
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Ketterman M., Madan A., Rodrigues S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalilus D.E.,  
RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
RT and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [3]  
RN IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR  
RP COMPLEX WITH NCOR2 AND HDAC3.  
RX PubMed=10809664;  
RA Guenther M.G., Lane W.S., Fischle W., Verdin E., Lazar M.A.,  
RA Shiekhattar R.;  
RT "A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-  
RT repeat protein linked to deafness.";  
RL Genes Dev. 14:1048-1057(2000).  
RN [4]  
RN COMPONENT OF THE N-COR COMPLEX WITH NCOR2 AND HDAC3.  
RX PubMed=10944117;  
RA Li J., Wang J., Wang J., Nawaz Z., Liu J.M., Qin J., Wong J.;  
RT "Both corepressor proteins SMRT and N-CoR exist in large protein  
RT complexes containing HDAC3.";  
RL EMBO J. 19:4342-4350(2000).  
RN [5]  
RN SUBUNIT OF A COMPLEX WITH UBE2D1, CACYBP, SIAH1 AND APC.  
RX PubMed=11389839;  
RA Matsuzawa S.-I., Reed J.C.;  
RT "Siah-1, SIP, and Ebi collaborate in a novel pathway for beta-catenin  
RT degradation linked to p53 responses.";  
RL Mol. Cell 7:915-926(2001).  
RN [6]  
RN COMPONENT OF THE N-COR COMPLEX WITH NCOR1, NCOR2, GPS2, TBL1R AND  
RP HDAC3.  
RX MEDLINE=21930350; PubMed=11931768;  
RA Zhang J., Kalkum M., Chait B.T., Roeder R.G.;  
RT "The N-CoR-HDAC3 nuclear receptor corepressor complex inhibits the JNK  
RT pathway through the integral subunit GPS2.";  
RL Mol. Cell 9:611-623(2002).  
RN [7]

RP COMPONENT OF THE N-COR COMPLEX WITH TBL1R; CORO2A AND HDAC3, AND  
RP HISTONE-BINDING.  
RX PubMed=12628926;  
RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,  
RA Wong J.;  
RT "Purification and functional characterization of the human N-CoR  
RT complex: the roles of HDAC3, TBL1 and TBL1R.";  
RL EMBO J. 22:1336-1346(2003).  
RN [8]  
RN FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.  
RX PubMed=14980219;  
RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;  
RT "A corepressor/coactivator exchange complex required for  
RT transcriptional activation by nuclear receptors and other regulated  
RT transcription factors.";  
RL Cell 116:511-526(2004).  
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the  
CC ubiquitin/19S proteasome complex to nuclear receptor-regulated  
CC transcription units. Plays an essential role in transcription  
CC activation mediated by nuclear receptors. Probably acts as  
CC integral component of corepressor complexes that mediates the  
CC recruitment of the 19S proteasome complex, leading to the  
CC subsequent proteasomal degradation of transcription repressor  
CC complexes, thereby allowing cofactor exchange.  
CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least  
CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.  
CC Component of a E3 ubiquitin ligase complex containing UBE2D1,  
CC SIAH1, CACYBP/SIP, SKP1A, APC and TBL1X. Probably part of other  
CC corepressor complexes, that do not contain NCOR1 and NCOR2.  
CC Interacts with histones H2B, H3a and H4.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- TISSUE SPECIFICITY: Ubiquitous.  
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and  
CC apparently displays the same function as component of ubiquitin E3  
CC ligase complexes (By similarity).  
CC -!- DISEASE: Defects in TBL1X may be a cause of late-onset  
CC sensorineural deafness (OASD) [MIM:300650]. OASD is an X-linked  
CC recessive disorder characterized by ocular albinism and  
CC progressive sensorineural hearing loss in the fourth and fifth  
CC decades of life. OASD may be caused by deletion of both GPR143/OAI  
CC and TBL1X adjacent genes; TBL1X defects possibly causing the  
CC hearing phenotype.  
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.  
CC -!- SIMILARITY: Contains 1 F-box-like domain.  
CC -!- SIMILARITY: Contains 1 Lish domain.  
CC -!- SIMILARITY: Contains 8 WD repeats.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; Y12781; CAA73319.1; ALT\_INIT.  
DR EMBL; BC032708; AAH32708.1; -  
DR EMBL; BC052304; AAH52304.1; ALT\_INIT.  
DR HSSP; P16649; IERJ. TBL1X.  
DR Genew; HGNC:11585; TBL1X.  
DR MIM; 300196; -  
DR MIM; 300650; -  
DR GO; GO:0007605; P:perception of sound; TAS.  
DR GO; GO:0007165; P:signal transduction; TAS.  
DR GO; GO:0007601; P:visual perception; TAS.  
DR InterPro; IPR006594; Lish.  
DR InterPro; IPR011047; Quin\_alc\_DH\_like.  
DR InterPro; IPR001680; WD40-  
DR Pfam; PF00400; WD40; 8.  
DR PRINTS; PR00320; GPROTEINRPT.  
DR ProDom; PD000018; WD40; 3.  
DR SMART; SM00667; Lish; 1.  
DR SMART; SM00320; WD40; 8.

```
DR PROSITE; PS00896; LISH; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00882; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Activator; Deafness; Nuclear protein; Repeat;
KW Transcription regulation; Ub1 conjugation pathway; WD repeat.
FT DOMAIN 4 36
FT DOMAIN 41 86
FT REPEAT 179 218
FT REPEAT 235 274
FT REPEAT 276 315
FT REPEAT 318 358
FT REPEAT 359 398
FT REPEAT 401 449
FT REPEAT 452 491
FT REPEAT 493 525
FT DOMAIN 108 124
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F6B9 CRC64;

Query Match          91.7%; Score 3355.2; DB 1; Length 526;
Best Local Similarity 86.2%; Pred. No. 3.1e-59;
Matches 456; Conservative 38; Mismatches 17; Indels 18; Gaps 10;

Qy 1 MSISDEVNLFVRYLQESGFSHSAFTGKSHISQSNINGALVPPAALISIIQKGLQYV 60
Db 1 MSITSDEVNLFVRYLQESGFSHSAFTGIESHISQSNINGTLVPPAALISILQKGLQYV 60

Qy 61 EAEVINEDEGTLFGRDIESLSLIDAVMPDVQTRQQAQYRDKLAAQQ--AAAAAAA-AA 117
Db 61 EAEISINEDGTVFGRDIESLSLIDAVMPDVQTRQQAQYRDKLAAQQASAAAAATAA 120

Qy 118 ASQ-----QG-S-----AKNGENTANGENGHTIANNHTDM-MEVGDGVIIPNKAVLVR 166
Db 121 ATAATTAGVSHQNPKNREAVNGENRAHSV-NNHA-KPMEDIDGEVIPSATVLR 178

Qy 167 GHESEVFIKAWNPVSDLLASGSDSTARIWNLSNSTSGSTQLVLRHCIREGGQDVPSNK 226
Db 179 GHESEVFIKAWNPVSDLLASGSDSTARIWNLSNENSGSTQLVLRHCIREGGHVDPSNK 238

Qy 227 DVTSLDWNSGTLTLAGSYGDFARITWKDGNLSTLQHGKPTPALKNWKNKGFILSAGV 286
Db 239 DVTSLDWNTGTLTLAGSYGDFARITWEDGNLSTLQHGKPTPALKNWKNKGYILSAGV 298

Qy 287 DKTTIWDATGTGAQKQPPHSPALPDVDWQSN--TFASCTSDMCIHVKLGODRPIKTF 345
Db 299 DKTTIWDATGTGAQKQPPHSPALPDVDWQ--NNTTFASCTSDMCIHVKLGODRPIKTF 357

Qy 346 QGHTNEVNAIKWDPGTGNLLASCDSDMTLKIWSMKQDNCVHDLQHNKEIYTIKWSPTGPG 405
Db 358 QGHTNEVNAIKWDPGMLLASCDSDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 417

Qy 406 TNNPNANMLASAFDSFTVLVDVDRGICHTLTKHQPYSVAFSPDGRYLAGSGFDKC 465
Db 418 TSNPNINMLASAFDSFTVLVDIERGVCTHTLTKHQPYSVAFSPDGRYLAGSGFDKC 477

Qy 466 VHIWNTOTGLVHSYRTGTGIFECVNAAGDKVCASASDGSVCVLDLRK 514
Db 478 VHIWNTUSGLVHSYRTGTGIFECVWNARGDKVCASASDGSVCVLDLRK 526

RESULT 7
ID TBL1 HUMAN
AC Q9BQ87; STANDARD; PRT; 522 AA.
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein TBL1Y (Transducin beta-like 1Y protein)
DE (Transducin-beta-like 1, Y-linked).
GN Name=TBL1Y; Synonyms=TEL1;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A., AND TISSUE SPECIFICITY
MEDLINE=22699787; PubMed=12815422; DOI=10.1038/nature01722;
Skaltsky H., Kuroda-Kawaguchi T., Minx P.J., Cordum H.S.,
Hiller L.W., Brown L.G., Repping S., Pyntikova T., Ali J., Bieri T.,
Chinwalla A., Delehaanty A., Delehaanty K., Du H., Fewell G.,
Fulton L., Fulton R., Graves T.A., Hou S.-F., Latrielle P.,
Leonard S., Mardis E., Maupin R., McPherson J., Miner T., Nash W.,
Nguyen C., Ozersky P., Pepin K., Rock S., Rohlfing T., Scott K.,
Schultz B., Strong C., Tin-Wollam A., Yang S.-P., Waterston R.H.,
Wilson R.K., Rozen S., Page D.C.;
"The male-specific region of the human Y chromosome is a mosaic of
discrete sequence classes.";
Nature 423:825-837(2003).
-!- FUNCTION: F-box-like protein involved in the recruitment of the
ubiquitin/19S proteasome complex to nuclear receptor-regulated
transcription units. Plays an essential role in transcription
activation mediated by nuclear receptors. Probably acts as
integral component of corepressor complexes that mediates the
recruitment of the 19S proteasome complex, leading to the
subsequent proteosomal degradation of transcription repressor
complexes, thereby allowing cofactor exchange (By similarity).
-!- SUBUNIT: Probable component of the N-CoR repressor complex and
some E3 ubiquitin ligase complex (By similarity).
-!- SUBCELLULAR LOCATION: Nuclear (By similarity).
-!- TISSUE SPECIFICITY: Fetal brain and prostate.
-!- DOMAIN: The F-box-like domain is related to the F-box domain, and
apparently displays the same function as component of ubiquitin E3
ligase complexes (By similarity).
-!- SIMILARITY: Belongs to the WD-repeat EBI family.
-!- SIMILARITY: Contains 1 F-box-like domain.
-!- SIMILARITY: Contains 1 Lish domain.
-!- SIMILARITY: Contains 8 WD repeats.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; AF332220; AAK13472.1; -.
EMBL; AF332221; AAK13473.1; -.
EMBL; AF332222; AAK13474.1; -.
HSP; P16649; IERJ.
MIM; 400033; -.
InterPro; IPR006594; Lish.
InterPro; IPR001680; WD40.
InterPro; IPR011046; WD40_like.
Pfam; PF00400; WD40; 8.
PRINTS; PR00320; GPROTEINBRPT.
ProDom; PD000018; WD40; 3.
SMART; SM00667; Lish; 1.
SMART; SM00320; WD40; 8.
PROSITE; PS00896; LISH; 1.
PROSITE; PS00678; WD REPEATS 1; 4.
PROSITE; PS00882; WD REPEATS 2; 6.
PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Activator; Nuclear protein; Repeat; Transcription regulation;
KW Ub1 conjugation pathway; WD repeat.
FT DOMAIN 4 36
FT DOMAIN 41 86
FT REPEAT 179 218
FT REPEAT 233 272
FT REPEAT 274 313
FT REPEAT 316 354
FT REPEAT 357 396
FT REPEAT 399 447
FT REPEAT 450 489
FT REPEAT 491 521
```



RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Aeshburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,  
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fletschmann W.,  
 RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.B., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195 (2000).  
 RN [3]  
 RN SEQUENCE FROM N.A.  
 RP STRAIN=Berkeley; Tissue=Embryo;  
 RP MEDLINE=22426066; PubMed=12537569;  
 RX Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,  
 RA George R.A., Guarin H., Konnoller B., Pacleb J.M., Park S., Wan K.H.,  
 RA Rubin G.M., Celniker S.E.;  
 RT "A *Drosophila* full-length cDNA resource."  
 RL Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8 (2002).  
 RN [4]  
 RN FUNCTION, INTERACTIONS WITH PHYL AND SINA, AND MUTANTS EBIC1 AND  
 RP EBIC3.  
 RX PubMed=11032805;  
 RX Boulton S.J., Brook A., Staehling-Hampton K., Heitzler P., Dyson N.;  
 RT "A role for Ebi in neuronal cell cycle control."  
 RL EMBO J. 19:5376-5386 (2000).  
 RN [5]  
 RN FUNCTION, AND INTERACTIONS WITH SNO; SMR AND SU(H).  
 RP PubMed=12230979;  
 RX Tsuda L., Nagaraj R., Zipursky S.L., Banerjee U.;  
 RA "An EGFR/Ebi/Sno pathway promotes delta expression by inactivating  
 RT Su(H)/SMRTER repression during inductive notch signaling."  
 RL Cell 110:625-637 (2002).  
 CC -1- FUNCTION: F-box-like component of E3 ubiquitin ligase complexes;  
 CC involved in R7 photoreceptor cell differentiation, cone cell  
 CC development and neuronal cell cycle control. E3 ubiquitin ligase  
 CC complexes mediate ubiquitination and subsequent proteasomal  
 CC degradation of target proteins. Required for specification of R7  
 CC photoreceptor cell fate in the eye by participating to the  
 CC ubiquitination and subsequent proteasomal degradation of Tramtrack  
 CC (ttk), a general inhibitor of photoreceptor differentiation.  
 CC Required to block the S phase entry in the peripheral nervous  
 CC system and central nervous system in a process that does not  
 CC involve the degradation of ttk. Involved in cone cell development

CC by preventing the transcriptional repression mediated by Su(H) on  
 CC DL, probably by participating to a E3 complex that contains sno  
 CC and mediates the ubiquitination and subsequent proteasomal  
 CC degradation of some component of the Su(H) repressor complex.  
 CC -1- SUBUNIT: Component of some E3 complex at least composed of sina,  
 CC ebi and phyl, required for the degradation of ttk. Probably forms  
 CC a E3 complex with sno, required for the degradation of some  
 CC component of the Su(H) repressor complex. Interacts with sno and  
 CC Su(H) and Smr.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Widely expressed both in embryos and larvae.  
 CC -1- DOMAIN: The F-box-like domain is related to the F-box domain and  
 CC apparently displays the same function as component of ubiquitin E3  
 CC ligase complexes.  
 CC -1- MISCELLANEOUS: 'Ebi' means 'shrimp' in Japanese.  
 CC -1- SIMILARITY: Belongs to the WD-repeat EBI family.  
 CC -1- SIMILARITY: Contains 1 F-box-like domain.  
 CC -1- SIMILARITY: Contains 1 Lish domain.  
 CC -1- SIMILARITY: Contains 8 WD repeats.  
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to a  
 CC frameshift in position 484.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: AF146345; AAD35017.1; --  
 CC EMBL: AB003589; AAF51501.1; --  
 CC EMBL: AF061326; AAL28874.1; ALT\_FRAME.  
 CC HSSP: Pf6649; 1ERJ.  
 CC FlyBase: FBgn0023444; ebi.  
 CC GO: GO:000074; P:regulation of cell cycle; IMP.  
 CC InterPro: IPR006594; Lish.  
 CC InterPro: IPR001680; WD40.  
 CC Pfam: PF00400; WD40; 3.  
 CC PRINTS: PR00320; GPROTEINRPT.  
 CC PRODOM: PD000018; WD40; 3.  
 CC SMART: SM00667; Lish; 1.  
 CC SMART: SM00320; WD40; 8.  
 CC PROSITE: PS50896; Lish; 1.  
 CC PROSITE: PS00678; WD\_REPEATS\_1; 3.  
 CC PROSITE: PS50082; WD\_REPEATS\_2; 6.  
 CC PROSITE: PS0294; WD\_REPEATS\_REGION; 1.  
 CC Cell cycle; Nuclear protein; Repeat; Ub1 conjugation pathway; Vision;  
 KW WD repeat.  
 FT DOMAIN 4 36 Lish.  
 FT REPEAT 41 84 F-box-like.  
 FT REPEAT 353 392 WD 1.  
 FT REPEAT 408 447 WD 2.  
 FT REPEAT 449 488 WD 3.  
 FT REPEAT 491 531 WD 4.  
 FT REPEAT 532 571 WD 5.  
 FT REPEAT 574 622 WD 6.  
 FT REPEAT 625 664 WD 7.  
 FT REPEAT 666 700 WD 8.  
 FT DOMAIN 206 323 Ser-rich.  
 FT MUTAGEN 16 16 L->Q: In ebiCC1; induces ectopic S phases  
 FT within the peripheral and central nervous  
 FT system.  
 FT C->Y: In ebi4; induces defects in R7 cell  
 FT development in the eye.  
 FT S->L: In ebiCC3; induces ectopic S phases  
 FT within the peripheral and central nervous  
 FT system.  
 SQ SEQUENCE 700 AA; 72387 MW; 28C6D8D07BB79FB7 CRC64;  
 Query Match 79.7%; Score 2917.1; DB 1; Length 700;  
 Best Local Similarity 57.4%; Pred. No. 1.2e-49;  
 Matches 411; Conservative 50; Mismatches 36; Indels 219; Gaps 58;

```
QY 1 MSISSEVNFLVYRYLQESGFSAFTFGIKSHISQSNINGALVPPAALISIIQKGLQV 60
D 1 MSPSSDEVNFLVYRYLQESGFSAVFGIESHISQSNINGALVPPAALLTIQKGLLT 60
QY 61 EAEVSIINEDGTLFD-CRPTESLSLIDAVMPDV-----VQTRQO-----AYRD---KLA-- 104
D 61 EVESWVGEDG---EVARPIEGLSLIDAVMPDVKPLPIVKT-EPGKPGAV-DSSAP-AGG 114
QY 105 -Q-----Q-----Q-A-----AA-A-----A 112
D 115 NQNNNAKPEIKIEPGTGVAGSAGCNKIAGSTGTSTPTDOSASEVDSSGNAANAGCTVA 174
QY 113 ---A-----A-----AA-ASQ-----Q-----GSAK--N-G-EN---T-- 130
D 175 GNGAGCAGNQAOSTGGSNSTSTPPAGGDLAAPGASQKKSQNSNEAGSSSGNAGNANATSTD 234
QY 131 -A-----NG-----EE-----N-----G-AH-----TI- 141
D 235 AASSTNGNSTSSSVQPTSLTTPAGGTVTSNPDAAAGGASTATGSKAPSGAVTIR 294
QY 142 --A--NN-----H-----T-DWMEV--DGD--VEIPNKAVVLRG 167
D 295 VQAQNNVQSGSSNAOSSAPSGTISSTSGGACTPAAL-VPMDIDENIPEISKARVLRG 353
QY 168 HESEVFICAWNPNVS-DLLASGSGDSTARIWNLSE-NSTSGSTQVLRLHCIREGGQDVPSN 225
D 354 HESEVFICAWNPN-SRDLASGSGDSTARIWMSDAN-T-NSNQLVLRHCLQKGAEPVSN 410
QY 226 KDVTSLDWSNBSGTLTATGSDGPARIW-TKDGMLASTLQGHKPIPALKNKK-GNFILS 283
D 411 KDVTSLDWCNDSGLLATGSDGVARIWKT-DGRLASTLQGHKPIPALKNW-KCGNYILS 468
QY 284 AGVDKTTIITDWAHTGE-AKQPFPHSAPALDWDQSNNTFASGSTD-MCHVCKLG--QD 339
D 469 AGVDKTTIITDWAHTGCT-QQFAFHSAPALDWDQNTQAFASGSTDQR-IHVCRLGVNE- 525
QY 340 RPIKTPQGHTEVNAIKWDPGTGNLLASCSDDMTLKIWSMKQD-NCVHDLQOHNKETIYTIK 398
D 526 -PIKTPKGTNEVNAIKWCPQGLLASCSDDMTLKIWSMNRDCC-HDLQAHSEKIYTIK 583
QY 399 WSPGTGCTNNPNANMLASGDSFSTVRLWDVDRGICIHITLTKHQBEPVYSAFSPDGRYLA 458
D 584 WSPGTGCTNNPNANMLASGDSFSTVRLWDVDRGICIHITLTKHQBEPVYSAFSPDGKHLA 643
QY 459 SSGFDCVCHVNTOTGALVHSYRGTCGIFEVWCNNAAGDKVGASASDGSCVCLDLRK 514
D 644 SSGFDCVCHVNTOTGALVHSYRGTCGIFEVWCNNAAGDKVGASASDGSCVCLDLRK 699

RESULT 10
Q96RW7 PRELIMINARY; PRT; 5636 AA.
AC Q96RW7
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hemicientin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Trent J.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156100; AAK68690.1; -.
DR HSSP; P07996; 1LSL.
DR GO; GO:0005524; F-ATP binding; IEA.
DR GO; GO:0005509; F-calcium ion binding; IEA.
DR GO; GO:0005021; F-vascular endothelial growth factor receptor. .; IEA.
DR GO; GO:0006468; P-protein amino acid phosphorylation; IEA.
DR InterPro; IPR000152; Asx_hydroxyl_S.
```

```
DR InterPro; IPR000875; Cecropin.
DR InterPro; IPR000742; EGF 2.
DR InterPro; IPR001881; EGF Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR006605; G2F.
DR InterPro; IPR009030; Grow_fac_recept.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003598; Ig_c2.
DR InterPro; IPR000169; Pept_cys_acsite.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR009134; VEGFR.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF07645; EGF CA; 8.
DR Pfam; PF07474; G2F; 1.
DR Pfam; PF00047; Ig; 44.
DR Pfam; PF00090; TSP 1; 6.
DR PRINTS; PR01832; VEGFRECEPTOR.
DR SMART; SM00179; EGF CA; 7.
DR SMART; SM00408; IGC2; 43.
DR SMART; SM00209; TSP1; 6.
DR SMART; SM00327; VWA; 1.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS00268; CECROPIN; UNKNOWN_1.
DR PROSITE; PS01186; EGF 2; 3.
DR PROSITE; PS00266; EGF 3; 5.
DR PROSITE; PS01187; EGF CA; 8.
DR PROSITE; PS00835; IG LIKE; 44.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.
DR PROSITE; PS00092; TSP1; 6.
KW EGF-like domain.
SQ SEQUENCE 5636 AA; 613673 MW; F00B319CED7B52C CRC64;

Query Match 71.5%; Score 2614.8; DB 2; Length 5636;
Best Local Similarity 9.5%; Pred.No. 1.4e-41;
Matches 436; Conservative 63; Mismatches 13; Indels 4072; Gaps 371;

QY 1 M-S-----I-----SS---D-----EV----- 8
D 1 MISWEVHTVPLFALLYSLLAQDASPOSIRAEPEPEGASTLAFVDFVTGSMYDDLQVVI 60
QY 9 -----NP-LV----- 12
D 61 EGASKILETSLKRPKRPFLNFALVPHDPEIGFVTTITDPKRFQVBLRELYVQGGDCPE 120
QY 13 -----Y-----R---Y---LQ----- 17
D 121 MSIGAIAKTALEISLPGSFYVFTDARSKDYRLTHEVLQLIQKQSQVFLTGDCDDRTH 180
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
D 181 IGYKVVEEIASTSSQGVFHLDDKKQVNEVLKQVNEAVQASKVHLLSTDHLEQAVNTWRIFP 240
QY 27 -----T-----TG-----RG-----I-----K 31
D 241 DPSLKEVTVLSGSPSMIEIRNPLGKLKKGFLHELLNIHNSAKVNVNKEPEAGMTVK 300
QY 32 ---S--H-----I-----S---Q----- 36
D 301 TSSSGRHVRITGLSTIDFRAGFSRKPTLDFKKTVSRPVQGIPTVYLLNTSGISTPARID 360
QY 37 -----S-----S-----NI-----N-----G----- 41
D 361 LLELLSISGSLKTIPIVKYYPHKKPYGIWNISDFVPPNEAFFLKVYTGDKDYLFORVSS 420
QY 42 -----A-----L-----VP-----P-----A 47
D 421 VSFSSIVPDAPKVTMPKPTPGYTLQPGQIPCSVDSLLPFTLSFVRNGVTLGVDQYLKESA 480
QY 48 -----A---L-----I-----S-----IIO-----KG----- 56
D 481 SVSLDIKVTLSDEGFYECIAVSSAGTGRAQTFDFVSEPPPIQVNNVTVTPGERAVLT 540
QY -57 -----L--Q-----Y-----V--E----- 61
```

Db 541 CLIIASVDYNTLWQRNDRVRLAEPARITRLANLSLEKSVKFNDAGEYHCWVSSEGSS 600  
QY 62 -A-----EVS-----NE-----68  
Db 601 AASVFLTVQBPPKVTVMKQNSFTGGSEVIMCSATGPKPKIAWTVNDMFIVGHRYRM 660  
QY 69 -DGTLF-----D-----G-----R-----P-----I--78  
Db 661 TSDGTLFIKNAAPKADAGIYCLASNSAGTDKQNSTLYIEAPKLMVVOSELLVALGDITV 720  
QY 79 -E--S-----L--S--LI-----DA-----V-----87  
Db 721 MECKTSGIPPPQKFKGDLRLPSTELIDPLGLLKIQETQDLQDAGDYTCVAINEAGR 780  
QY 88 -----M-----P-DV-----VQ-----T-----94  
Db 781 ATGKITLDVGSPPVFIQEPADVMEIGSNVTLFCYVQGYPEPTIKWRRLDNMFI SRPFS 840  
QY 95 -----R-----Q--Q--Q--Q--AA-----109  
Db 841 VSSISQRLTGALFILNLWASDKGYICEAENQFKIQSETTVTGLVAPLIGISPVAN 900  
QY 96 ---QQ-----A-----Y-----R-----D--KL--103  
Db 901 VIEGQQLTLPCTLLAGNPERRWIKNSAMLLQNPYITVRS DGLHIERVOLQDGGEYTC 960  
QY 104 ---A-----Q--Q--Q--Q--AA-----109  
Db 961 VASNAGTNNKTTVVVHVHVLPTIQHQQLLSTIEGIVTLPCKASGNPKPSVWSKKGEL 1020  
QY 110 ---AAA---A-A-----A--AA--A-----118  
Db 1021 ISTSSAKFSAGAGSLVWVSPEGESEGYVCTAINTAGYAKRKVQLTVYVRVFGDLRG 1080  
QY 119 -SQ-----Q-----Q-----GS-----123  
Db 1081 LSQDKPVEISVLAGEEVLTPCEVKSLLPPIITWAKETQLISPFSPRHTLPSPGSMKITET 1140  
QY 124 -----A-----K-----N-----126  
Db 1141 RTSDSGMYLCVATNIAGNVQAVKLNHVHPKIQRGPKHLKVQGVQDVIPCAQAQTPLP 1200  
QY 127 -----GE--N--T--A-----N--G-----133  
Db 1201 VITWSKGSMTLVDGHHVSNPDGTLSDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260  
QY 134 -----E-----E--N-----GA-----H-----139  
Db 1261 PPTVEDLEPPYNTTFQERVANQRIEFPKPAKGTPTIKWLHNGRELTGREPGISILEEG 1320  
QY 140 T--IA-----N-----N--HT-----DM-----148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHPVPVVKDKBQVSNVSVLLNQLTNL 1380  
QY 149 -----M-----E-----150  
Db 1381 FCEVEGTPSPILMYKQNVQVTESSITQTVNNGKILKLFRATPDAGRYSCKAINAGTS 1440  
QY 151 ---VD-----G-----DV--E-----I-----P-----PN--160  
Db 1441 QKYFNIDVLVPPTIIGTNFPKSVNLNRDVALEQVKGTPFPDIHWFKDGKPLFLGDPN 1500  
QY 161 -----K-A-----V-----VL--165  
Db 1501 VELLDRGQVLHLKNARNNDKGRYQCTVSNAAQKQAKDIKTIYIPPSIKGNVTDISVL 1560  
QY 166 -----RG-----H-----168  
Db 1561 INSLIKLECKTRGLPMPAITWYKDGQPIWSSQALYIDKQYLHPIPAQVSDSATYCHV 1620  
QY 169 -----E-S-----E-----VFI-----C--A-----W-----177

Db 1621 ANVAGTAESFHVVDVYVPPWIEGNLATPLNKQVVIHSLTLECNAAGNPSPILTLWKDGV 1680  
QY 178 -----N-----P-V--SD--182  
Db 1681 PVKANDNFRIAGGKKLEIMSAQSIDRGQYICVATSVAGEKEIKYEVDVLPVPAIEGGDE 1740  
QY 183 -----LL-----ASGS-----G-----D-----S--191  
Db 1741 TSYFIVVNNLLELDCHVTGSPPTIMLWKDQGLIDRDFKILLNGRKLVIQAQVSNT 1800  
QY 192 -----TA-----R-----I-W--196  
Db 1801 GLYRCMAANTAGDHKEFEVTHVPTTIKSSGSLSERVWVKYKVALQCIANGIPNSITW 1860  
QY 197 -----NL-----S-----S--199  
Db 1861 LKDDQPWTAQGNLKIQSSGRVLQIAKTLLDAGRYTCVATNAAGETOQHILHVHPPS 1920  
QY 200 -E-----N-----S-----T-----SGST-----207  
Db 1921 LEDAGMLNETVLVSNPVLQECKAAGNPVPTIWKDNCLLSGSTMTFLNRQIIDIES 1980  
QY 208 -Q-----L--V-----LR-H-----C--I--215  
Db 1981 AQISDAGIYKCVAINASAGATELFYSLQVHVAFPSISGSNNMVAVVNNVPRLECEARGIPA 2040  
QY 216 -----R-----E--G-----G--Q--D--221  
Db 2041 PSUTLWKDGPVSSFSNGLOVLSGGRIALATSTQISDTGRTYTCVAVNAAGEKORDILRV 2100  
QY 222 -VPSN-----KD-----VT--S-L--231  
Db 2101 YVPPNIMBEEQNVSVLISOAVELLQSDAI PPPTTLTWLKDGHPLKKPGLSISENSVLK 2160  
QY 232 -----D-----W--N--S-----EG--TLL--240  
Db 2161 IEDAQVQDTGRTYTCVATNVAGTEKKNVNNVWPPNIGSGDELTQLTVEGNLISLCE 2220  
QY 241 AT-----GS--Y--D-----G-----F-----A--249  
Db 2221 SSGIPPNLIWKKGSPVLTDSMGRVRLSGRQLQISAEKSDAALYSCVASNVAGTAK 2280  
QY 250 -----R-----I-----WTKDG-----256  
Db 2281 KEYNLQVIRPTITNSGSHPTLIVTRGKISLECEVOGIPPTVTWMDGHPLIKAKGV 2340  
QY 257 -----NL-AS-----T--L-----GQH-----265  
Db 2341 EILDEGHILQKNHIVSDTGRYCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400  
QY 266 -K-----G-PI-----FALK--W-----274  
Db 2401 EKNVSLTCEASGIPLPSTWTF--KDGWPVLSNSVRILSGGRMLRLMQTTMEDAQYTC 2458  
QY 275 -----N-----K--K-----GN-----279  
Db 2459 VVRNAGEERKIFGLSVLPVPHIVGNTLEDVKVKEQSVTLTCEVTGNPVPVETWHKDG 2518  
QY 280 -----F--I-----L--S--AGV--286  
Db 2519 QPLODEAHIIISGGRFLQITNVQPHGTGRYTCCLASSPAGHKRSFSLNVFVSPTIAGVG 2578  
QY 287 -D--K--T-----T--I--W-----293  
Db 2579 SDGNPDVTILNSPTSLVCEAYSPPATITWFKDGTPLESNRNRILPGRTLQILNAQ 2638  
QY 294 -D-----A-----H-----TG-----E--299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIIPINKGLWPGLSPEKVIKVNNTLLECE 2698  
QY 300 A-----K-----Q-----Q-----303  
Db 2699 AVAIPASLSWYKDGOPKLSDDHVNIAANGHTLQIKEAQISDTGRTYTCVASNIAGEDEL 2758

304 F-----P-F-----H- 307  
2759 FDVNIQVPPSFQKLEIGNMLDTCRNGEAKDVIINNPIISLYCETNAAPPTLTWYKQHP 2818  
308 --SA-----P-----A-----LDV-----D 315  
2819 LTSKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878  
316 -----WO-----SN----- 319  
2879 LPEVTVLVNKSALIECLSSGSPAPRNSWKDQOPLEDHKKFLSNGRILQILINTQITD 2938  
320 -----NT-----FAS--C-----S 326  
2939 IGRYVCVAENTAGSACKYFNLNVHVPSPVIGPKSENLTVVVNFISITCEVSGFPDLS 2998  
327 -----T-----D-----MCI----- 331  
2999 WLKNXQPIKLTNTLIVPGGRTLQIIIRAKVSDGGEYTCIAINXAGESKKFSLTVVPPS 3058  
332 --H-----V--C-----K-----L--GO----- 338  
3059 IKHDSLSLVNVRGTSVLSCESNVAPPPVITWYKGRMITESTHVEILADQMLHI 3118  
339 -----D-----R-----P-ik----- 343  
3119 KKAESDVTGOYVCRAINAGRDDKNFHLNVVYVPPSIEGPEREVIVETISNPVTLTCDATG 3178  
344 -----T-----F-----O-----G-----H-----T----- 349  
3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGSKLOIARSQSDSGNYTCIASNMEGKAQKY 3238  
350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
3239 FLSIQVPPSVAGAEIPSDVSVLGENVELVCNANGIPTLIQWLKDGKPIASGETERIRV 3298  
361 --G--NLL-A--S-----C--S-----D-----D-----MTL-- 373  
3299 SANGSTLNIYGALTSDTGKYTCVATWPAGEEDRIFNLNVVYVTTIRGNKDEAKLTYVD 3358  
374 --K-----I-W-----SM-----KQ--D-----NCV----- 384  
3359 TSINIECRXTGTPPPQINMLKNGLPLPLSLSHIRLLAAGQVIRVRAQVSDVAVYTCASN 3418  
385 -----H-----D-----L-----Q----- 388  
3419 RAGVDNKHYNLQVAFPPNMDNSMGTEITVLKGSSTSMACITDGTAPASNAWLRDGQPLG 3478  
389 -----Q-----Q-----H-N-----KB 393  
3479 LDAHVTSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFLKVLPPHINGSEHEE 3538  
394 I-----YT-----I--K-W----- 399  
3539 ISVINNPLELTCTIASGIPAPKMTWMDKGRPLQDQVOTLGGGEVLRISTAQVEDTGRY 3598  
400 -----SPTG-----P--GT-----N----- 407  
3599 TCLASSPAGDDKEYILVRVHVPPIAGTDEPRDITVLRNRQVLTCECKSDAAPPVPTWLR 3658  
408 N-----P-----N-----ANLM-LAS--A--S--F----- 420  
3659 NGERLOATPRVRLTSGRYLIQINNADLGDNTANTYCVASNIAGTKTREFILTIVNVPNIKG 3718  
421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433  
3719 GPQSVILLNKSTVLECIAGVPTPRITWRKQGAVLGNHARYSILENGFLHIQSAHVD 3778  
434 -----C-----I-----H-----TL-----T----- 439  
3779 TGRYLCMATNAAGTDRRRDLQVHVPPSIAPGPTNMTVNVVQTTLACEATGIPKPSINW 3838

440 -K--H-----Q-E-----P-V-----Y----- 446  
3839 RKNGLHLLNDQNSRYLLSSGLVITISPSVDDTATYECTVTNAGDDKRTVDTLVQVPP 3898  
447 SVA-----F-S-----P-D----- 453  
3899 STADEPTDFLVTKHAPAVITCTASGVPFPIHWTKNGIRLLPRGDGYRILSSGAIEILAT 3958  
454 -----GRY-----L--ASG--S-- 461  
3959 QLNHAGRYTCVARNAGSAHRHVTLVHHEPPVIOPOPSSELHVLNPNPILLPCEATGTPSP 4018  
462 F-----D-----KCV----- 466  
4019 FITWQKEGINVNTSGRHHAVLPSSGGLQISRAVREDAGTVMCAQNAGTALGKILNVQV 4078  
467 -----H-----I-W----- 469  
4079 PPVISPHLKEYVIANDKPITLSCADGLPPPDITWHKDGRAIVESIRQVLSGSLQIAF 4138  
470 -----N-----T-----Q----- 472  
4139 VQGDAGHYTCMAANVAGSSSTSTKLTVHVPPRIRSTEGHYTVNENSQAILPCVADGPT 4198  
473 -----T-----G-----A-----L-VH 478  
4199 PAINKKDNVLLANLLGKYTABPYGELILENVVLEDSGFVTCVANNAAAGSDHTVSLTVH 4258  
479 -----S-----Y-----R 481  
4259 VLPTTFELPGDVLNKGEOQLRSLCKATGIPKLTWTFNNNIIPAHFDSVNGHSELVIER 4318  
482 -----GT-----G-----IF-----E-----V--C----- 490  
4319 VSKEDSGTVCTAENSVMGFVKATGFVYKBPVFKGDYPSNWTIEPLGGNAILNCEVKGDP 4378  
491 -----WN-----A-----AGD----- 496  
4379 TPTIQNRKGVDEISHRIRQLNGSLAIYGTWNEDAGDYTCVATNEAGVVERMSLTQ 4438  
497 -----K-V--G-----A-----SAS----- 503  
4439 SPPITLEPVTETVAGGKIILNCQATGPQPTITTSRQGHSGISWDDRVNLSNNSLYIA 4498  
504 -----D-----GSV-----C-V----- 509  
4499 DAQKEDTSEFCVARNLMGSLVVRVPIVQVHGGFGSQNSAWRACSVTCCKGIQKRSLCN 4558  
510 --L-----DL--R-----K 514  
4559 QPLPANGKPKCGSDLEMCNQNK 4582

## RESULT 11

QWZ42  
ID QWZ42 PRELIMINARY; PRT; 34350 AA.  
AC QWZ42;  
DT 01-MAR-2002 (TREMELrel. 20, Created)  
DT 01-MAR-2002 (TREMELrel. 20, Last sequence update)  
DT 01-MAR-2004 (TREMELrel. 26, Last annotation update)  
DE Titin.  
GN Name=TTN;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20309627; PubMed=10850961;  
RA Freiburg A., Trombitas K., Hell W., Cazorla O., Fougereuse P.,  
RA Centner T., Kolmerer B., Witt C., Beckmann J.S., Gregorio C.C.,  
RA Granzier H., Labelt S.;  
RT "Series of exon-skipping events in the elastic spring region of titin





Db 4381 AAPVIRKIEBLEVALGHLAKFTCEIQSAPNRFQWFKAGREIYESDKCSIRSKYISSL 4440  
QY 134 E-----E-----N----- 136  
Db 4441 EILRTQVDCGEYTCASNEYGVSTATLVTVEAYPTFLSRPKSLTTFVGKAAFICT 4500  
QY 137 -----GA-----H-----TI-----A-N----- 143  
Db 4501 VTGTPVETIWDGGAALSPNWKLSAENKHILELSNLTIODRGVYSCASKNKFCADI 4560  
QY 144 -----N-----H----- 145  
Db 4561 COAELIIDKPHFIKELEPVQSAINKVHLEQWDEDRKVTVTWSDKGQKLPKGDYKIC 4620  
QY 146 -----T----- 148  
Db 4621 FEDKIATLEIPLAKLKDGYVCTASNEAGSSCSATVTVREPPSFVKKVDPSYMLPGE 4680  
QY 149 -----M-----E-----V-D-G----- 153  
Db 4681 SARLHCKLKGSPVIOVTFKNNKELSESNTVMYFVNSEAILDITDVKVEDSDSYCEAV 4740  
QY 154 -DV-----EI-----PP----- 159  
Db 4741 NDVGSDSCSTEIVIKEPPSPFKTLEPADIVRGTNALLQCEVSGTGPFSEISWFKDKQIRS 4800  
QY 160 -----N-----K----- 161  
Db 4801 SKYRLFQSKSLVLEIFSPNSADVGECVWANEVKGCCMATHLLKBPPTFVKVDDL 4860  
QY 162 -----AV-----VL----- 165  
Db 4861 IALGGQVTLQAAVRGSEPISTWMMKQEVIREDEGKIKMSFNGVAVLIIPDVQISFGGK 4920  
QY 166 -----R----- 166  
Db 4921 YTCLAENEAGSQTSGELIVKEPAKIERAELIQVTAGDPATLEYTVAGTPELKPKWKYKD 4980  
QY 167 G-----H-----E-S-EV-----F-----I-- 174  
Db 4981 GRPLVASKKYRISFKNNVAQLKPYSAELHDGQYTFEISNEVGSSCETTFVLDRDIAP 5040  
QY 175 -----C-A-----W----- 177  
Db 5041 PFTKPLRNVDSVNGTCRLDCKTAGSLPMRVSHFQKGEIAASDRYRIAPVEGTASLEII 5100  
QY 178 -----N-----P-V-----S-D----- 182  
Db 5101 RVDMDAGNFTCRATNSVGSKDSSGALIVQEPSPFVTKPGSKDVLPGSAVCLKSTFGST 5160  
QY 183 -L-----L-----A-----SG----- 187  
Db 5161 PLTIRWFGKNKELVSGGCVYITKEALESSLELYLVKTSQYTKCVSNVAGGVECSANL 5220  
QY 188 -----S-----GD----- 191  
Db 5221 FVKEPATFVEKLEPSQLLKKGDATQOLACKVTGTPPIKITWFANDREIKESSKHRMSFVES 5280  
QY 192 TA--R----- 194  
Db 5281 TAVLRITDVGIEDSGEYMCQAQNEAGSDHCSSIVIVKESPYFTKEPKPIEVLKEYDVMLL 5340  
QY 195 -----I-W-N-----L-S-----E-----N----- 201  
Db 5341 AEVAGTPPFEITWFKDNTILRSGRKYKFTQIDHLVSLQILKFAADAGEYQCRVTNEVS 5400  
QY 202 -----STS-----GST--Q-----L--V--L-----R 212  
Db 5401 STCSARVTLREPPSPFKTSTSSLRGGTAQFATLKGSLPITVTWTKDSDEITEDDNIR 5460  
QY 213 -----H-----C-----I-R-----E----- 217

Db 5461 MTFENNVASLYLSGIEVKHDKGYVCOAKNDAGIQRCQSALLSVKEPATITEEAVSIDVTOG 5520  
QY 218 -----G-----GO----- 220  
Db 5521 DPATLQVFGSGYKEITAKWFKDQBELTGLSKYKISVTDVTISILKIISTEKQDSGEYTFEV 5580  
QY 221 --DV-----PS-NK-----D-VT--S-LD-----W----- 233  
Db 5581 QNDVGRSSCKARINVLDLIIPPSFTKKLKMDSIKGSFIDLECIAGSHFISIQWPKDDQ 5640  
QY 234 -----N-----S-E-----GTL--AT--G-----S-----Y----- 245  
Db 5641 EISASEKYKSFHDNTAFLEISQLEGTDSGTVTCSATNKAGHNQCSGHLTVKEPPYFVEK 5700  
QY 246 --D-----G-----F-----AR-IW-----T-----K-- 254  
Db 5701 POSQDVPNTRVOLKALVGTAPMTIKWFDKNKELHSGAARSVWKDDTSTLSLELFAKAT 5760  
QY 255 D-G-----N-----L-----AST----- 261  
Db 5761 DSGTYICQLSNDVGTATSKATLTVKEPPQFIKKPSPVLVRNGQSTTFECQITGTPKIRV 5820  
QY 262 --L-G-----Q-H-----K-----G----- 267  
Db 5821 SWYLDGNEITAIQKHGISPIDGLATFQISGARVENSQTVVCEARNDAGTASCSELKVKE 5880  
QY 268 -PIF--ALK-----W--N-----KK----- 277  
Db 5881 PPTFIRELKPVEVKYSDVELECEVTGTPPPEVTWLNKNNREIRSSKKYTLTDRVSVFNLH 5940  
QY 278 -----G-----N-----FI-----L--SA----- 284  
Db 5941 ITKCDPSDGTGQCVISNEGSCSTRVALKEPPSPFIKKIENTTTVLKSSATFQSTVAG 6000  
QY 285 -----G-----GV----- 286  
Db 6001 SPPISITWLKQDILDEDDNVYISFVDSVATLQIRSVNDHSGRYTQAKNBSGVRCYA 6060  
QY 287 ----- 286  
Db 6061 FLLVQEPQIVKAKSVQVTEKDPMTLECVAGTPELKVKWLDKGQIVPSRYFSMSFN 6120  
QY 287 -----D-----K-----T-----T----- 290  
Db 6121 NVASFRIQSMKQDSQYTFKVENDEGSSSCDAYLRVLQDNIPPSFTKKLTQMDKVLGSS 6180  
QY 291 I-----I--W-D-----A-----H-----T----- 297  
Db 6181 IHMECKVSGSLPISAQWFKDQKGEISTSAKYRLVCHERSVSVLEVNLEEDTANYTCKVSN 6240  
QY 298 -----G-----E-----AK--QO-----P----- 306  
Db 6241 VAGDDACSGILTVKEPPSPFLVPGRQQAIPDSTVEFKAILKGTPPPKIKWFKDDVBLVSG 6300  
QY 307 -----H----- 307  
Db 6301 PKCFITGLESTSFNLNYSVDASKTQYTCVTNDVSDSCTTMLLVTEPPKFKVKLEASK 6360  
QY 308 -----S-----A-----PA-----LD--V-----D 315  
Db 6361 IVKAGDSSRLECKIAGSPEIRVVMFRNEHELPAQKRYMTFIDSVAVIQNNLSTEDSGD 6420  
QY 316 ----- 316  
Db 6421 FICEAQNPAAGTSCSTKVIKPEPPVFPSSPPPIVETLKNAEVSLCELSGTTPPEVVMYKD 6480  
QY 317 --Q-----S--N-----N-----T-----F-- 322  
Db 6481 KQQLASSKKYKIAKNFHTSIHLNVDTSDIGEYHCKAQNVEGSDTCVCTVKLKEPPRFV 6540  
QY 323 -----AS----- 324  
Db 6541 SKLNSLTVVAGBPAELQASIEGAQPIFQWLKEKEVIRESENIRITFVENVATLOFAKA 6600

QY 325 ---C---STD---M---C--- 330  
Db 6601 EPANAGYICQIKNDGGMEENWATLMVLEPAVIVEKAGPMVTVGETCTTLECKVAGTPEL 6660  
QY 331 -I---H---V---C--- 334  
Db 6661 SVEMKDGKLTSSQKHKFSFYNKISSILRILSVRQDAGTYTFQVQNVGKSSCTAVVDV 6720  
QY 335 ---K---LG--- 337  
Db 6721 SDRVPPSFTRRLKNTGVLGASCILECKVAGSSPISVWPFHEKTIKIVSGAKYQTFESDN 6780  
QY 338 ---Q---D---R---P--- 341  
Db 6781 VCTIQLNSLSDMGNTCVANVAGSDECAVLTVOBPSPFVKEPPELVLPKGNVTF 6840  
QY 342 -I---K---T---F---Q---G--- 347  
Db 6841 SVIRGTPPFKVNFRGARELVKGRCNYPEDTVAELELFNIDISQSGEYTCVVSNNAQ 6900  
QY 348 ---H---T---N--- 350  
Db 6901 ASCTRLFVKEPAFLKRLSDHSVEPGKSHILESTYTGTLPISTVTKDGFNITSEKN 6960  
QY 351 ---EV---N---A---IK--- 356  
Db 6961 IVITEKTCILEILNSTKRDAGQSCIEIENAGRDVCGALVSTLEPPYFVTELEPLEAAG 7020  
QY 357 ---W---D---PT---G--- 361  
Db 7021 DSVSLQOVAGTPEITVSWYKGDPTKLRPTPEYRTYFTNNVATLVFNKVNINDSGEYTC 7080  
QY 362 -N---L---LA---SC---S--- 368  
Db 7081 ENSIGTASSKTVFRIQERQLPPSPFARQLKDIEQTVGLPVLTLTCLRLNGSAPIQVCWYRDGV 7140  
QY 369 ---DD---MTLKI--- 375  
Db 7141 LLRDDENLQTSFVDNVATLKLQTLDSHQYSCSASNPLGTASSARLTAREPKKSPFF 7200  
QY 376 ---WS--- 377  
Db 7201 DIKPSIDVIAGESADPECHVTGAPMRTITWSKONKEIRPGNVTITCVGNTPHLRILKV 7260  
QY 378 ---M---KQ--- 380  
Db 7261 KGDSGYTQCOATNDVGKMCQAQLSVKPPKFVKKLEASKVAKQGESIQLECKISGSPE 7320  
QY 381 ---D---N---C-VH--- 385  
Db 7321 IKVSWFENDSELHESWKYNNSFINVALLTINEASAEKSDGYICEAHNGVGDASCSTALT 7380  
QY 386 ---D---LQ---Q--- 389  
Db 7381 VKAPPVFTQKSPVGVALKGSDVILQCEISGTPPEVWVKDRQVRNKKFKITSKHFD 7440  
QY 390 ---H---NKE---I---Y---IK--- 398  
Db 7441 SLHLNLEASDVGEYHCKATNEVGSDFCSKSVKPEPPRFVKLSPTSLIGDAVELRAI 7500  
QY 399 ---W---SP--- 401  
Db 7501 VEGFQPSVWVKDRGEVIRESENTRISFDNIATLQSGPEASNSGKYICQIKNDAGMR 7560  
QY 402 ---TG-P---GT--- 406  
Db 7561 ECSAVLTVPARIIEKPEMTVTGNPFALCECVTGTPELSAKWFKDGRSADSKHHI 7620  
QY 407 ---N---N---P---NA--- 411  
Db 7621 TFINKVASLKIPCAEMSDKGLYSFEVKNVSGKNCTVSVHVSRIVPPSPFIRKLKDVNAI 7680

QY 412 ---NL-ML--- 415  
Db 7681 LGASVLECRVSGSAPISVGVFQDGNIEVSGPKQCSFSFSENVCTLNLSLLEPDSGTGIYTC 7740  
QY 416 ---A---SA---SF---DS---T---VR---L---W--- 427  
Db 7741 VAANVAGSDSCSAVLTVQBPPEQTPDSVEVLPGMSLTFTSVIRGTPPKVKWFKGSRE 7800  
QY 428 --- 427  
Db 7801 LVPGESCNISLEDFVTELELFEVQPLESGDYSCLVNDAGSCTTHLFFVKBPATFVKRL 7860  
QY 428 -D---V---D---R--- 431  
Db 7861 ADFSVETGSPVLEATYTGTPPISVSWIKDEYLIQSERCSTMTWTKSTILELESTIED 7920  
QY 432 ---G---IC---I---H---TL---T--- 439  
Db 7921 YAQYSLIENAGODICEALVLEPPYFIEPLEHVEAIVGEPATLQCKVDGTFEIRISW 7980  
QY 440 -K-H---Q---E--- 443  
Db 7981 YKEHTKLRSAAYKMOFKNNVASLVINKVDHSDVGEYSCKADNSVGAVASSAVLVIKERK 8040  
QY 444 -P---V---YSVAF---S-P--- 452  
Db 8041 LPFFPARKLKVHETLGPVAFECRINGSPELQVSWYKGVLLKDDANLOTSPVHNVATL 8100  
QY 453 ---D---GRY-L-AS---G---S---FD---KC 465  
Db 8101 QILQTOQSHIGQYNCASNPLGTASSAKLILSEHEVPPFDLKPVSVDLALGESGTFKC 8160  
QY 466 VH---I-W---NT---Q-T---G- 474  
Db 8161 -HVTGAPIKITWAKDNREIRPGNYKMTLVENTATLTVLKVKGDAQGYTCVASNIAG 8219  
QY 475 ---A---L--- 476  
Db 8220 DSCSAHLGVQEPFRFIKLEPSRIVKQDEFTRYECKIGSGPEIKVLWYKDETEIQESSKF 8279  
QY 477 ---V---H---S-Y---R--- 481  
Db 8280 RMSFVDSVAVLEMHNLVSDSGDYTCEAHNAAGSASSSTLKVKEPPIFRKKPHPIETLK 8339  
QY 482 GT---G---G---I---P---E--- 488  
Db 8340 GADVHLECLQGTTPPHVSWYKDKRELRSKKYKIMSENFLLTSIHILNVDAADIGYQCK 8399  
QY 489 ---V---C--- 490  
Db 8400 ATNDVGSDFCVGSIALKAPRFVKVLLSDISTVVGVKEVQLQTTIEGAEPISVVWFKDKGEI 8459  
QY 491 ---W---N---A-AG---D--- 496  
Db 8460 VRESNIMWISYENATLQSRVPEANAGKYTCQIKNDAGMOECFATLSVLEPATIVEKP 8519  
QY 497 ---KV-G---A---S---A-SD 504  
Db 8520 ESIKVTTGTCTLECTVAGTPELSTKWKFGDKELSDNKKYSFFNKNVSGLKIIINVAPSD 8579  
QY 505 -G-S---V---CV-L---D---L-RK 514  
Db 8580 SGVYSFEVQNPVGKDSCTASLQVSDRTVPPSPTRK 8614

RESULT 12  
AAP49011 PRELIMINARY; PRT; 7073 AA.  
ID AAP49011  
AC AAP49011;  
DT 02-MAR-2004 (TREMELrel. 27, Created)  
DT 02-MAR-2004 (TREMELrel. 27, Last sequence update)  
DT 02-MAR-2004 (TREMELrel. 27, Last annotation update)  
DE Orfiab polyprotein.



QY 167 -----CHE-----S-----E----- 171  
Db 2036 KEVIEDVKTTEVGVNVLKPSDEGVKVTQELGHEDLMAAYVENTSITIKKNELSLALG 2095  
QY 172 -----VF----- 173  
Db 2096 LKTIATHGIAINSPWSKILAYVKPFLGQAAITTSNCAKLAQRFVNNMYPVFTLLFQ 2155  
QY 174 -----I-----C-----A-W----- 177  
Db 2156 LCTFTKSTNSRIRASLPTTIKNSVKSAKLCLDAGINYKSPKSLFTIAMWLLLSI 2215  
QY 178 -----N----- 178  
Db 2216 CLGSLICVTAAGVLLSNFGAPSYCNGVRELYLNSNVTTMDFCBGFPCISCLGLDSL 2275  
QY 179 -----P-----S-----DL-----LA-----S-----G----- 189  
Db 2276 DSYPALETIOVTISSYKLDLTLGLAAEWLAWMLFTKPYLLGLSALMQVFFGYFASHF 2335  
QY 190 -----D-----ST-----A 193  
Db 2336 ISNSLWMTIISIVQMAPSAMVRMYIFFASFYIWKSYVHMDGCTSSCTMCMYKRNRA 2395  
QY 194 -----I-----WN-----ISEN-----ST-----SGST----- 207  
Db 2396 TRVECTIVNGMKRSFYVYANGRGCFKTHWNCL--NCDTCTGSTFTISDEVARDLSLQ 2453  
QY 208 -----Q-----L-----V-----L-----RH----- 213  
Db 2454 FKRPINPTDOSSYIVDSVAVNGALHLYFDKAGQKYERHPLSHFVNLDLNRANTKGS 2513  
QY 214 -----C-----I----- 215  
Db 2514 PINVIVFGSKCDESASVYSQMLCQPIILLDQALVSDVGDSTEVSRMFDAYVD 2573  
QY 216 -----R----- 216  
Db 2574 TFSATFVPMKULKALVATAHSELAKGVALDGLVSTFVSAAQGVDTDTVDKDVIECLK 2633  
QY 217 -----E-----G----- 220  
Db 2634 LSHHSDLEVTDGSCNNFMTYNKVENMTPRDLGACIDCNARHINAQVAKSHNVSLIWNVK 2693  
QY 221 D----- 221  
Db 2694 DYNLSBQLRQIRSAAKNNIPRLTCAITRQVNVNITTKISLGGKIVSTCFKLMLKA 2753  
QY 222 -----V-----P-----S-----N-----K-----D-----VT-----S-----LD----- 232  
Db 2754 TLLCVLAALVCYIYMPVHTLSIHGTYNEIIGYKAIQDGVTRDIIISTDDCFANKHAGFDA 2813  
QY 233 W-----N-----S-----E-----GT----- 238  
Db 2814 WFSQGSYKNDKSCPVVAALITFEIGFIVPGLPGTVLRAINGDFLHFLPRFVSAGNIC 2873  
QY 239 -----L-----LA-----T-----G-----SY----- 245  
Db 2874 YTPSKLIEYDFATSACVLAECTIFKDMGKVPYCYDNTLLEGSISYSELRPDTRYVL 2933  
QY 246 -----DG-----F-----A-----R-----I-----W----- 252  
Db 2934 MDGSIIOFPNTYLEGSVRVVTTFDAEYCRHGTCERSEVGICLSTSGRWVLNNEHYRALSG 2993  
QY 253 -----TK-----D----- 255  
Db 2994 VFCGVDMNLIANITFPIVQPVGALDVSASVAVAGGIITAILVTCAAYFMKFRFVGEYNH 3053  
QY 256 -----G-----N-----LA-----S----- 260  
Db 3054 VVAANALLFLMSFTILCLVPAYSLPGYVFLYLYLTFTFTNDVSFLAHLQWFMFSPIV 3113

QY 261 -----T-----L-----G----- 263  
Db 3114 PFWITAIYVFCISLKHCHWFFNNYLRKRVFNGVFTSTFEBAALCTFLLNKEMYLKRSE 3173  
QY 264 -----Q-----H-----K-----G-----P----- 268  
Db 3174 TLLPLTQYRNRYLALYNKYKYPFGALDITTSYREAAACCHLAKALNDFNSGADLYYQPOTS 3233  
QY 269 I-----F-----AL-----K-----W----- 274  
Db 3234 ITSAVLQSGFRKMAFPGSGKVEGCMQVTCGTTTLNGLWLDLDDTVYCPRHVICTAEDMLNPN 3293  
QY 275 -----N-----K----- 276  
Db 3294 YEDLLIRKSNHSLVQAGNVQLRVIGHSMQNCCLRLKVDTSNPKTPKYKRVRIQPGQTS 3353  
QY 277 -----KG-----N-----F-----I-----L-----S-----A 284  
Db 3354 VLACYNGSPSGVYQCAMPNHTIKGSLNGSCGSGVFNIDYDCVSFCYMHMELPTGVHA 3413  
QY 285 G-----VDK-----TTI-----I-----W-----D----- 294  
Db 3414 GTDLGKFGYFPVDRQTAQAAGTDTITLNLVLAWLYAAVINGDRMFLNFTTLNDPNLV 3473  
QY 295 A-----H-----TG-----E----- 299  
Db 3474 AMKYNVEPLTQDHDVILGFLPSAQTGIADVLMCAALKELLQNGMNGRTILGSTILEDEFTP 3533  
QY 300 -----AK-----Q-----O-----P-----PP----- 306  
Db 3534 FQVVRQCGVTFQKFKIVKGTTHHMLLTFLTLLVQSTQWSLFFPVYENAFLEPTL 3593  
QY 307 -----H-----S-----A-----PA----- 311  
Db 3594 GIMATACAMLLVKKHAFCLFLPLSLATVAYFNNVMPASVWVRIMTWELADTSLSG 3653  
QY 312 -----L-----D-----V-----D-----W-----Q----- 317  
Db 3654 YRLKDCVMYASALVLLILMTARTVYDAAARRVMTLMNVITLVYKVYGNALDQAIMWAL 3713  
QY 318 -----SN-----NT----- 321  
Db 3714 VISVTSNYSGVTTIMFLARAIVFCVEYVYPLLTGNTLQCIIMLVYCFGLYCCCYFGL 3773  
QY 322 FASC-----ST-----D----- 328  
Db 3774 F-----CLLNRYFRLTLGVYDLVSTQEFYNNMSOGLLPKSSIDAFKLNKILGIGGKPCIK 3831  
QY 329 -----M-----CI-----H----- 332  
Db 3832 VATVQSKMSDVKCTS VVLLSVLQQLRVSESSKLAQCQVQLHNDILLAKDTTEAFKMWSL 3891  
QY 333 -----V-----CK-----L-----G----- 337  
Db 3892 LSVLLSMOAGVDTNRLCEEMLDNRATLOAIASEFSLPSYAAAYATAQAEYEQAVANGDSE 3951  
QY 338 -----Q----- 338  
Db 3952 VVLKUKKSLNVAKSEFDRDAAMQKLEKMAQAMQYKQARSEDKRAKVTSAMQTMFLF 4011  
QY 339 -----D-----R-----P-----I-----K-----TF----- 345  
Db 4012 TMLRKLNDALNNIINNARDGCVPLNIIPLTTAAKLMVVVDPDYGTGKTCGNTFTYASA 4071  
QY 346 -----Q-----CH-----T-----NE-----V----- 352  
Db 4072 LWEIQVVDADSKHVQLSEINMDNSPNLAWPLIVTALRANSVKLQNNELSPVALQMSC 4131  
QY 353 -----NA-----I-----KW-----D-----P----- 359  
Db 4132 AAGTTQACTDDNALAYNNKGRFVALLSDHQDKWARFPKSDGTGTIYTELEBPCR 4191  
QY 360 -----T-----G-----NL-----LA-----S-----CS----- 368

Db 4192 FVTDTPKGPVKYLYFIKGLNNLRGMVLSLAATVRLQAGNATEVPAHSTVLSFCFAV 4251  
QY 369 D---D---M---T---  
Db 4252 DPAKAYDLASGGQPTNCVMKLTHTGTGQAITVTPANMDQESFGGASCCLYCRCHI 4311  
QY 373 ---LK---I---W---S---M---  
Db 4312 DHPNPKGFCDLKGYQIPTTCANDPVGFTLRNTVCTVGMWKGCGSCDQLREPLMQSA 4371  
QY 379 ---Q---K---  
Db 4372 DASTFLNRVCGVSAARLTPCGTGTSTVDVYRAPDIYNEKVAGFAKFLKTNCCRFQEKDEE 4431  
QY 381 ---D---N---C---VHDL---Q---  
Db 4432 GNLLDSYFVVVKRHTMSNYQHEETIYNLVKDCPAVAHDFPKFRVGDGMVPHISRQLTKY 4491  
QY 389 ---OH---N---KEI---  
Db 4492 TMADLVVLRHFDGEGCDTLKEILVTYNCDDDYFNKCDWYDFVENPDILLRVYANLGERV 4551  
QY 395 ---Y---  
Db 4552 ROSLLKTVQCDAMRDAGIVGLTLDNQLNGWYDFGFEQVAPGCGVPVDSYYSLLM 4611  
QY 396 ---T---IKW---  
Db 4612 PILTTRALAESHMADLAKPLIKWDLKYDFTTEERLCLFDRYFKYWDQTYHPNCINCL 4671  
QY 400 ---S---PT---GP---  
Db 4672 DDCILHCANFNLFSTVPTPTSGPLVRKI FVDGVPFVSTGYHPRELGVHVNQDNLH 4731  
QY 405 ---G---TNN---P---  
Db 4732 SSRLSPKELLVYAADPAMHAASGNLLDKRTTCFSAALTNNVAFQTVKPGNFKDFYDF 4791  
QY 410 ---NA---NL---M---L---  
Db 4792 AVSKGPFKEGSSVELKHFHFAQGNAAISDYDYRYNLPNTMCDIROLLFVVEVDKYFDC 4851  
QY 416 ---A---SASF---DS---  
Db 4852 YDGGCINANOVINNLDKSAGFPFNKWKARLYYDSMSYEDQDALFAYTKBNVIPITILM 4911  
QY 423 ---TV---R---L---W---  
Db 4912 NLKYAISAKNRARTVAGVSTCTMTNRQFHQKLLSIAATRGATVIGTSKFYGGWHNML 4971  
QY 428 ---DV---DR---  
Db 4972 KTVYSDVETPLMGWDYPKCDRAMPNMLRIMASLVLRKHNTCCNLSHRFYRLANECAV 5031  
QY 432 ---G---IC---  
Db 5032 LSEWVCGGSLYKPGTSSGDATTAYANSVFNICQAVTANVALLSTDGNKIADKYVN 5091  
QY 436 ---H---TIT---KH---  
Db 5092 LQHLRYECLYRNRDVEHDFEYFAYLRKHFSMMILSDDAVVCYNSNAAQGLVASIKNF 5151  
QY 442 ---Q---E---P---VY---S-V---A---  
Db 5152 KAVLYYQNNVFMSEAKWTETDLTKGPHBFCSQHTMLVKQDDYVYLPDPSRILGAGC 5211  
QY 450 ---F-S---P---D---G---  
Db 5212 FVDDIVKDTGLMIEFVSLAIDAYPLTKHPNQEYADVFLHYLYQYLRKLHDELTHMLDM 5271  
QY 455 ---RY---L---A---S---GS---F---

Db 5272 YSVMLTNDNTSRWEPFEYEAMYTPHTVLOAVGACVLCNSQTSRLRCGACIRRPFLCKCC 5331  
QY 463 -D---K---C---V---H---  
Db 5332 YDHVISTSHKLVLSNPPYVCNAPGCDVDTQLYLGMSYYCKSHKPPISFPLCANGQVF 5391  
QY 468 ---I---W---NT---  
Db 5392 GLYKNTCVGSDNVDFNAIATCDWTNAGDYILANTCTERLKLFAAETLKATEETFKLSYG 5451  
QY 472 ---Q---T---G-A---  
Db 5452 IATVREVLSDRELHLHSGWEVGKPRPLNRNRYVFTGYRTNKNKQVIGEYTFEKGDDYDAVV 5511  
QY 476 ---L---V---H---S---Y-R---  
Db 5512 YRGTTYKLVNGVDYFVLTSTHTWPLSAPTLVPOEHVVRITGLYPTLNIISDFSSNVANYQ 5571  
QY 482 ---GTG---G---  
Db 5572 KVGMMQKYSTLOGPPGPGTKSHFAIGLALYYPSARIVYTACSHAADVADALCEKALKYLPIDKC 5631  
QY 486 ---I---F---E---  
Db 5632 SRIIPARARVECFDKFKNSTLEQVYVCTVNALPETTADIVVFEISMATNYDLSVNNAR 5691  
QY 489 ---VC---  
Db 5692 LRAKHVYVIGDPAQLPAPRTLLTKGLEPEYFNSVCRMLMKTIGDPMFLGTCRCPCPAEIVD 5751  
QY 491 ---W---  
Db 5752 TVSALVYDNDKLKAHKDKSAQCFKMFYKVI THDVSSAINRPOIGVVVREFLTRNPAMRKAV 5811  
QY 492 ---NA-A---G---D---KV 498  
Db 5812 FISPYNQNAVASKILGLPTQTVDSSQSGSEYDVIPTTETAHSCNVNRFNVAITRAKI 5871  
QY 499 G---AS-A---S---D---G---SV 507  
Db 5872 GILCIMSORDLYDKLOFTSLEIPRRNVATLQENVTGLFKDCSKIITGLHPTQAPTHLSV 5931  
QY 508 ---CV---L-D---LR---K 514  
Db 5932 DIKFKTEGLCVDIPGIPKDMTYRRLISMWGF 5963  
RESULT 13  
LRIB MOUSE  
ID LRIB MOUSE STANDARD; PRT; 4599 AA.  
AC Q9J118; Q8BZD3; Q8BZM7;  
DT 05-JUL-2004 (Rel. 44, Created)  
DT 05-JUL-2004 (Rel. 44, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Low-density lipoprotein receptor-related protein 1B precursor (Low-density lipoprotein receptor-related protein-deleted in tumor) (LRP-DIT).  
DE Name=Lrp1b; Synonyms=Lrpdit;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]\_TaxID=10090;  
RP SEQUENCE FROM N.A.  
RA Yakichkin S., Lisitsyn N.;  
RT "The structure of the mouse homologue of the human candidate tumor suppressor gene LRP1B/LRP-DIT";  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 3770-4599 FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,

RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tonaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schirali L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,  
RA Blake J.A., Bradt V., Brusci V., Clothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Glissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustinich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,  
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wyshaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sakaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RA Birney E., Hayashizaki Y.,  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RL 60,770 full-length cDNAs";  
RL Nature 420:563-573(2002).  
CC -|- FUNCTION: Potential cell surface proteins that bind and  
CC internalize ligands in the process of receptor-mediated  
CC endocytosis.  
CC -|- SUBCELLULAR LOCATION: Type I membrane protein (Potential).  
CC -|- SIMILARITY: Belongs to the LDLR family.  
CC -|- SIMILARITY: Contains 17 EGF-like domains.  
CC -|- SIMILARITY: Contains 32 LDL-receptor class A domains.  
CC -|- SIMILARITY: Contains 25 LDL-receptor class B domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AF270884; AAF81558.1; -;  
DR EMBL; AK035795; BAC29188.1; -;  
DR EMBL; AK034122; BAC28594.1; -;  
DR HSSP; Q07954; 1CR8.  
DR MGD; MGI-2151136; Lrp1b.  
DR InterPro; IPR011044; Amine\_DH\_B\_like.  
DR InterPro; IPR001881; EGF\_CA.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR009030; Grow\_fac\_recept.  
DR InterPro; IPR002172; LDL\_receptor\_A.  
DR InterPro; IPR000033; Ldl\_receptor\_rep.  
DR Pfam; PF00008; EGF; 15.  
DR Pfam; PF00057; Ldl\_recept\_a; 32.  
DR Pfam; PF00058; Ldl\_recept\_b; 33.  
DR PRINTS; PR00261; LDLRECEPTOR.  
DR SMART; SM00179; EGF\_CA; 4.  
DR SMART; SM00192; LDLa; 32.  
DR SMART; SM00135; LY; 36.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 4.  
DR PROSITE; PS00022; EGF\_1; 5.  
DR PROSITE; PS01186; EGF\_2; 9.  
DR PROSITE; PS50026; EGF\_3; 9.  
DR PROSITE; PS01187; EGF\_CA; 3.  
DR PROSITE; PS01209; LDLRA\_1; 27.  
DR PROSITE; PS50068; LDLRA\_2; 32.  
KW Calcium-binding; EGF-like domain; Endocytosis; Glycoprotein; Receptor;  
KW Repeat; Signal; Transmembrane.  
FT SIGNAL 1 20 Potential.

CHAIN	21	4599	Low-density lipoprotein receptor-related protein 1B.
FT DOMAIN	21	4444	Extracellular (Potential).
FT TRANSMEM	4445	4467	Potential.
FT DOMAIN	4468	4599	Cytoplasmic (Potential).
FT DOMAIN	31	70	LDL-receptor class A 1.
FT DOMAIN	76	114	LDL-receptor class A 2.
FT DOMAIN	120	154	EGF-like 1.
FT DOMAIN	155	194	EGF-like 2, calcium-binding (Potential).
FT DOMAIN	195	336	LDL-receptor class B 1.
FT DOMAIN	338	380	LDL-receptor class B 2.
FT DOMAIN	382	424	LDL-receptor class B 3.
FT DOMAIN	568	609	LDL-receptor class B 4.
FT DOMAIN	611	655	LDL-receptor class B 5.
FT DOMAIN	657	705	LDL-receptor class B 6.
FT DOMAIN	707	754	LDL-receptor class B 7.
FT DOMAIN	794	834	EGF-like 3.
FT DOMAIN	844	882	LDL-receptor class A 3.
FT DOMAIN	885	923	LDL-receptor class A 4.
FT DOMAIN	926	963	LDL-receptor class A 5.
FT DOMAIN	966	1003	LDL-receptor class A 6.
FT DOMAIN	1005	1043	LDL-receptor class A 7.
FT DOMAIN	1052	1089	LDL-receptor class A 8.
FT DOMAIN	1094	1132	LDL-receptor class A 9.
FT DOMAIN	1135	1174	LDL-receptor class A 10.
FT DOMAIN	1172	1213	EGF-like 4.
FT DOMAIN	1300	1345	LDL-receptor class B 8.
FT DOMAIN	1347	1388	LDL-receptor class B 9.
FT DOMAIN	1390	1435	LDL-receptor class B 10.
FT DOMAIN	1437	1480	LDL-receptor class B 11.
FT DOMAIN	1481	1521	LDL-receptor class B 12.
FT DOMAIN	1527	1570	EGF-like 5.
FT DOMAIN	1618	1659	LDL-receptor class B 13.
FT DOMAIN	1661	1703	LDL-receptor class B 14.
FT DOMAIN	1705	1743	LDL-receptor class B 15.
FT DOMAIN	1745	1788	LDL-receptor class B 16.
FT DOMAIN	1834	1875	EGF-like 6.
FT DOMAIN	1922	1963	LDL-receptor class B 17.
FT DOMAIN	1965	2006	LDL-receptor class B 18.
FT DOMAIN	2008	2050	LDL-receptor class B 19.
FT DOMAIN	2052	2094	LDL-receptor class B 20.
FT DOMAIN	2143	2183	EGF-like 7.
FT DOMAIN	2239	2278	LDL-receptor class B 21.
FT DOMAIN	2281	2328	LDL-receptor class B 22.
FT DOMAIN	2330	2373	LDL-receptor class B 23.
FT DOMAIN	2375	2416	LDL-receptor class B 24.
FT DOMAIN	2417	2458	LDL-receptor class B 25.
FT DOMAIN	2509	2548	LDL-receptor class A 11.
FT DOMAIN	2551	2587	LDL-receptor class A 12.
FT DOMAIN	2590	2626	LDL-receptor class A 13.
FT DOMAIN	2629	2675	LDL-receptor class A 14.
FT DOMAIN	2681	2717	LDL-receptor class A 15.
FT DOMAIN	2719	2757	LDL-receptor class A 16.
FT DOMAIN	2760	2800	LDL-receptor class A 17.
FT DOMAIN	2804	2841	LDL-receptor class A 18.
FT DOMAIN	2844	2885	LDL-receptor class A 19.
FT DOMAIN	2890	2926	LDL-receptor class A 20.
FT DOMAIN	2926	2967	EGF-like 8.
FT DOMAIN	2968	3008	EGF-like 9.
FT DOMAIN	3055	3097	LDL-receptor class B 26.
FT DOMAIN	3099	3140	LDL-receptor class B 27.
FT DOMAIN	3142	3184	LDL-receptor class B 28.
FT DOMAIN	3186	3223	LDL-receptor class B 29.
FT DOMAIN	3225	3266	LDL-receptor class B 30.
FT DOMAIN	3274	3314	EGF-like 10.
FT DOMAIN	3316	3353	LDL-receptor class A 21.
FT DOMAIN	3356	3392	LDL-receptor class A 22.
FT DOMAIN	3395	3432	LDL-receptor class A 23.
FT DOMAIN	3435	3472	LDL-receptor class A 24.
FT DOMAIN	3475	3511	LDL-receptor class A 25.
FT DOMAIN	3514	3550	LDL-receptor class A 26.
FT DOMAIN	3552	3588	LDL-receptor class A 27.
FT DOMAIN	3593	3629	LDL-receptor class A 28.

FT	DOMAIN	3631	3668	LDL-receptor class A 29.	
FT	DOMAIN	3673	3711	LDL-receptor class A 30.	
FT	DOMAIN	3714	3752	LDL-receptor class A 31.	
FT	DOMAIN	3761	3797	LDL-receptor class A 32.	
FT	DOMAIN	3801	3843	EGF-like 11.	
FT	DOMAIN	3844	3881	EGF-like 12.	
FT	DOMAIN	3981	4037	LDL-receptor class B 31.	
FT	DOMAIN	4039	4081	LDL-receptor class B 32.	
FT	DOMAIN	4083	4126	LDL-receptor class B 33.	
FT	DOMAIN	4213	4249	EGF-like 13.	
FT	DOMAIN	4249	4290	EGF-like 14.	
FT	DOMAIN	4286	4326	EGF-like 15.	
FT	DOMAIN	4321	4363	EGF-like 16.	
FT	DOMAIN	4392	4430	EGF-like 17.	
FT	DISULFID	32	45	By similarity.	
FT	DISULFID	39	58	By similarity.	
FT	DISULFID	52	69	By similarity.	
FT	DISULFID	77	90	By similarity.	
FT	DISULFID	84	103	By similarity.	
FT	DISULFID	97	113	By similarity.	
FT	DISULFID	159	169	By similarity.	
FT	DISULFID	165	178	By similarity.	
FT	DISULFID	180	193	By similarity.	
FT	DISULFID	798	809	By similarity.	
FT	DISULFID	805	818	By similarity.	
FT	DISULFID	820	833	By similarity.	
FT	DISULFID	845	857	By similarity.	
FT	DISULFID	852	870	By similarity.	
FT	DISULFID	864	881	By similarity.	
FT	DISULFID	886	898	By similarity.	

Query Match  
Best Local Similarity 71.3%; Score 2610.5; DB 1; Length 4599;  
Matches 434; Conservative 56; Mismatches 16; Indels 4085; Gaps 365;

QY	1	MS	-----I-----S-----	4
Db	1	MS	QLLLAILTLGSLPNAEVLIVGANOQDHLCDPGEFLCHDHVTCVSQWLCDGDPDCPD	60
QY	5	-SD-	-----EV-----N-----P-----	10
Db	61	QSD	ESLDTCPPEVEIKCPLNHIAHGSSACVHLSKLCNGVVDGDFDEGGHCQELLPS	120
QY	11	-L-	-----V-----Y-----R-----Y-L-Q-----	17
Db	121	QQL	NCFKAMVNAFCYCEDGFEVAEDGRSCKDQDECSIYICISQTKNTYGSVACSC	180
QY	18	-ESGF-	-----S-----H-----SAF-T-----F-GTK-----S-----H-----	33
Db	181	VE	-GYIMQSDNRSCVKHEPTDKAPMLLISSLETIELFYINGSKMTLSSANRNEITHLD	239
QY	34	-----	-----I-S-QS-----NI-----	39
Db	240	FIY	EMICWIESRESSNLKCGOITKAGRLTDORIINSLSQSPQVQEQMAFDWLTRNIYF	299
QY	40	-----	-----NG-----A-L-----A-L-----VP-----	45
Db	300	VHV	SDRIFVFCNGSVCVTLIESELNPKAIADPIAGKLPFTDGNVPKIERCDLGM	359
QY	46	-----	-----P-A-AL-----I-----SIQKGLQ-----Y-59	
Db	360	NR	TRIVSKAEQPSALALDNLRLVYVDLYDVGVDYQGNRHTIVQ-GRQVKHLYG	418
QY	60	-V-E-	-----AEVS-----IN-----	68
Db	419	ITV	FEDLYA-TSSDNFIIRNFRNGTDIHSIKMESARGINTYQKRTQPTVRSHACEV	477
QY	69	D	-----G-----TL-----F-----DGR-----	76
Db	478	DAY	CMGGGSHICLLSSSVKTRCTCRCTGFNMGSDGRSCKRPKNELFLFYKGRPGIVRG	537
QY	77	-----	-----PIE-----SL-----S-LI-----	84

Db	538	MDL	NTKIADCEMIPBENLVNPRALDPFAEANYIYFADTTTFLIGRQKIDGTERETILKOD	597
QY	85	-D-	-----AV-----M-----P-----	89
Db	598	LON	VEGIAVDWIGNNLYWTNDGHRKTINVARLEKASQSRKTLLEGSMHPRAIIVDPNG	657
QY	90	-----	-----D-V-----V-Q-----	93
Db	658	MM	WTWDKEDKIDDSVGRIEKAMWDGVRQVFTSKMLPENGTLTDFHTSTLYWCDAIYD	717
QY	94	-----	-----T-R-----Q-----	96
Db	718	HIE	KVFLNGTHRKVVYSGKELNHPFGLSHHGNYVFTDYMNGSIFQLDLMTNEVTLRHE	777
QY	97	-----	-----QAV--R-----D-K-----LA-----	104
Db	778	RAP	LFGLQIYDRKQOGDMCRINNGCGTLCIAIPAGRVACADNQLLDENGTCTTFNP	837
QY	105	-----	-----Q-----Q-----Q-----Q-----	107
Db	838	EE	IRPHICKPGEFRCKNKHCIQARWKCDGDDCLDGSDEDSVTCFNHSCPDQDFKQNNR	897
QY	108	-----	-----A-----AA-----AA-----AA-----AA-----	116
Db	898	CIP	KWLCDGANDCGSNEDESNQTCRTARTQADQFSCGNRCIPTAWLCDREDDCGQTD	957
QY	117	-AAS-	-----Q-----Q-----Q-----Q-----	121
Db	958	EV	ASCEFTPCBLPTQFIKSGRCISNKHCHDTHDDCGDRSDEVCVHSCLDQDFRCSGR	1017
QY	122	-----	-----GS-A-----KN-G-----E-N-T-----A-----N-----GE-	134
Db	1018	CIP	GHWACDGDNDGDFSDETHINCKEARSAGCIGNFCQPCPGNCIPDLWRCDGK	1077
QY	135	-----	-----E-NGA-----H-----T-----I-----	141
Db	1078	DC	EDSGEKGCNGTIRLCHDKTKFCRSTGRCINNAWVGDDVDCEDQDEEDCDSFLCG	1137
QY	142	-----	-----AN-----NH-----	145
Db	1138	PP	KPCANDTSVCLQPEKLCNGRKDCPDGSDGDLCDCESLNNGCSNHCSSVVGRIVC	1197
QY	146	-----	-----T-----D-----MM-----E-----VD-	152
Db	1198	SC	PEGHQKKNRTCEIVDYCASHLRCSQVCEQKHVKSCVEGWALGTGDSCTSVDS	1257
QY	153	-----	-----GD-----	154
Db	1258	FE	AFIIFSIRHEIRIDLHKGDYSLIVPGLRNTIALDFHNSQLLYWTDVVEDRIYRGKL	1317
QY	155	-----	-----VE-----I-----	157
Db	1318	SS	GGVSAEVVVEHGLATPEGLTVDWIAAGNIYWIDSNLDQIEVSKLDGSLRATLAGAM	1377
QY	158	-----	-----P-----N-----K-----	161
Db	1378	EH	PRALDPYGLFWTDWANDANPRIESASMSGAGRTIYKDMKTGAMPNGLTVDHFER	1437
QY	162	-----	-----AV-----VLRGHE-----SEV-----	172
Db	1438	RIV	WDARSDAIYSAFYDGTNNMIEIRGHEYLSPFAVSLYGVVYVTDWRTNTLAKANK	1497
QY	173	-----	-----F-----I-----CA--	176
Db	1498	WT	QNVSVIQTSAOFPDLOIYHPSRQPOAPNPCAANEGRGPCSHLCLINHNRSACACP	1557
QY	177	-----	-----W-----NP-----VS-----D-----	182
Db	1558	HL	KLSSDKKTCYEMKKFLLYARSEIRGVDDIDNPVYNEITAFVDPDIDDAVIDDASE	1617
QY	183	-----	-----LL-----A-----SG-----S-G-----D-----S-----T-----	192
Db	1618	ER	LYWTDIKTQITRAFINGTGLTISRDIQSIIRGLAVDWSRNLVWISSEFETQINV	1677

193 AR-----I-----W-----N-----197  
1678 ARLOGLKTSIIHIDKPCQLAAHPVRKLYWTGNTINMANMDGNSKILFQKQEPVG 1737  
198 LS-----S-----T-----SG-----S-----T-----207  
1738 LSIYVENKLYWISSNGTINRCNLNLEVIEMKEELTKATALTIMDKLWADONL 1797  
208 OL-----VLR-----H-----213  
1798 AOLGTCKNRDGRNPSILRNKTSIGVVMKVDYDKEAQOQNSQVNNNGCSQLCLPTSETTR 1857  
214 -----C-----IR-----216  
1858 TCMCTVGYLQKRNMSQGIIESFLMYSVHEGIRGIPLEPRDKVDALMPISGAAPAVGIDF 1917  
217 -----EG-----218  
1918 HAENDTIYTDMLNKLISRAKRDQTKWEDVVTNGLGRVEGLAVDWDIAGNIYWTDDHGFNLI 1977  
219 -----GO-----D-----V 222  
1978 EVARLNGSPRYIIISQGLDOPRSIAVHPEKGFLEWTEWQVPCIGKARLDGSEKVMIVSV 2037  
223 ---P---S---NK---D---V---T---S---LD---232  
2038 GITWPGISIDYENKLYWCDARSJKIERIDLTGANREVLGSGNVDLFSVAVFGAIYI 2097  
233 W---N---S---E---GT---238  
2098 WSDRAHANGSVRRGHKNKDATETVMTRTGLGNLKEIKIPNRVREKGTNVCAKENGCGQOL 2157  
239 -----LLA---T---G-----S---Y---245  
2158 CLYRGNRRTCACAHGYLAGDVTCLRHGYLLYSGRTILKSILHSDETNLNSPVRYEN 2217  
246 -----D---G---F---A---R---I---251  
2218 PNYFNIIALAFDYNORREGNRIIFYSDAHFNIQLIKNWEDRQVIVENVGSVEGLAYH 2277  
252 -----W---TK---D---G-----NL-----258  
2278 RAWDTLYWTSSTSSITRHTVQTRPCAIDREAVITWSEDDHPHVLADCEQNLFWTNW 2337  
259 -----A---ST---LG-----Q---264  
2338 NEQHPSIMRATLTGKNHVVVSTDLTPNGLTIDHRAEKLYFSDGSLGKIERCEYDGSQR 2397  
265 H---K---GP-----IF-ALKW-----NK-----K-----277  
2398 HVIVKSGGTFLSLAVYDSYIFWS-DWGRRAILRSNYKTGGETKILRSDIPHQPMGLIAV 2456  
278 -----G-----280  
2457 ANDTNSCELSPCALLGGCHDLCLLTPDGRVNCRCGRVLLANNRCVTNKSNCNIYSEP 2516  
281 -----I-----L-----S---AG-----V-----286  
2517 ECGNGDCVYVLTCDGIPHCKDKSDEKLLYCENRSCRSRSGFKPCYNRRCPVHGKLCBDGTD 2576  
287 -----D---K---T-----T---I-----291  
2577 CGDSSDELDCKVSTCSTVEFCADGTICIPRSARCNQNMDCSDASDEKGCNNTDCTHYKL 2636  
292 -----I-----W-----D-----AH-----296  
2637 GVKSTGFI RCNSTSLCVLPWSICDGSNDGCDYSDLKPCPVONKHCBENYFCGSGRCIL 2696  
297 ---T---G---E-----A-----K-----301  
2697 NTWVCDGQKDCBDGLDLHCDSSCSMNQFACSVKCKISKHWICDGEDDCGDSLDESISIC 2756

302 -----Q-----Q-----F-----304  
2757 GAVTCAADMPSCGSHACVPHWLCDGERDCPDGSDLSAGCAPNNTCDENAFMCHNKV 2816  
305 ---P---F---H-----S-----AP-----ALD-----VD---WQ-----S 318  
2817 CIPKQFVCHDDDDCGSDSDFLQCGYRQCPBEFRCA-DGRCLVNTLWQCDGDFDCPDSS 2875  
319 ---N-----NT---FASC-----STD-MC-I-----HV---C---334  
2876 DEAPINPRCSRABHSNCSFFM-CKNGRCIPS-DGLCDIRDDCGSDSDETCHINECLSK 2933  
335 KL-G--OD-R--PI-----KT-----P-----O-----G-----347  
2934 KISGCSQCDLPSYKCKWPGFQLKDGKTCVDIDECSSGPPCQOCINTYGVTKYCHC 2993  
348 -----H-----T-----N-----350  
2994 AEGYETOPDNPNGCRSLSDPEPLIADQHEIRKISTDGSNYTLLKQGLNNVIALDFDYR 3053  
351 -E-----V-NA-----I-K---W-D-----358  
3054 EEFYIWDSSRPNGSRINRMCLNGSDIKVHNTAVENALAVDWIGKLYWSDTEKRIIEV 3113  
359 -----PT-----GNLL-ASC-----367  
3114 SKLNGLYPTVLYSKRLKFPRLSLDPAGNLYWIDCEYPHIGRVGMDGTNQSVIETKI 3173  
368 S-----D-----M-----TL-K-I-W-----376  
3174 SRPMALTIDVNHRLYADENHIEFSNMDGSHRHKVPNQDIPGVIALTLTFEDYIYWDGK 3233  
377 ---S---MK-----Q-D-----N-----C-----383  
3234 TKLSRVHKTSGADRLSLINSHAITDIQVHYSYRQPDVSKHLTVNNGGCHLCLLPG 3293  
384 -VH-----D---L-----387  
3294 KHTCACTNFYLAADNRCLNCTASQPRCKTDKIPFWWKCDTVDDCGSDSDEDDCP 3353  
388 ---Q-----Q-----H-----NKE-I---394  
3354 EFKCQPRFQCGTGLCALPAFICDGENDCGSDSDELNCDTHVCLAQFCKTKNKKCIPVN 3413  
395 -----Y-----T---I-K-W---SP-----401  
3414 LRNGQDDCGDEBEKDCPENSCSPDYFOCKTKKHCISKLWVCDDEDDPCADASEANCDK 3473  
402 -T-GP-----G-----T-N---N---P-----409  
3474 KTCGPHEFOCKNNNCIPDHWRCNQDNCSDNDEDNCKPQTCTLDPLCSNGDCVSSRW 3533  
410 -----N-----A-----N-----412  
3534 CDGEFDCADGSDKNCETSCSKDQFQCSNGQCLSAKWCKDGHEDCKYGEKNCPEAPV 3593  
413 -----LMLA-----SNS-----F-----420  
3594 CSSESVMASGGLSASLAKNGEPCDVGSDSDMDCVIECKEDQFQCKNKAYCIPIRWLCD 3653  
421 -----D-S---T-----V-R-----L-W-----D---V---429  
3654 GIYDCVDSDEFTCGRGGSI CRDDEFLNNLSLKLHFWVCDGEDDCGSDSDEAPMCKVF 3713  
430 -----DR-----GI-----433  
3714 LCPPTPRYCRNDRICLOLEKICNGINDCGSDSDEEHCGLSKLSPCKCKDEFTCSNRN 3773  
434 CI-----HT-----437  
3774 CIPMELQCSLDDCGSDSDEOGLKTPTEHTCENNGNPGCDDAYCNQIKTSVFCRCKPGF 3833  
438 -----L-T-----K-----440



```

Db 3834 QRNMKGRECADNELLFGICSHCLNTRGYSKVCQDNQFEKNNSCIAKSEDAQALYIA 3893
QY 441 -----HQ-----E-----P--V-Y-----446
Db 3894 NDTDLGFFVFFNYSGHQIISHVHNSRITGMDVHYQRNVIVWSTQFNEGGIFYKMDA 3953
QY 447 -----S-----VA-----FS-----451
Db 3954 REKQANSGLICPEFKRPDIAVDWAGNVVTDHSMHWFYTYHTWTSLSRYNSVQGL 4013
QY 452 -----P-----DG-----454
Db 4014 NGPNCRTLLTNMAGEPYAIAVNPGRGMVYTWVIGDHSIHBEAAMDGTLRVLVQKNLQRP 4073
QY 455 -----R-Y-----LA--S--GS-----461
Db 4074 TGLTVDFHGRITWADFELSIIGSLVDGSPVVSUSKOGLLPHRPHIDVFEDYIYGAGP 4133
QY 462 -----F-----DK-----464
Db 4134 KNGIFRVQKFGHGSVEVLALGVDKTKSILVSHRYKQLNLPNCPCLDLSCDFLCLNPSGAT 4193
QY 465 CV-----H-----I-----W-----N--470
Db 4194 CICPEGYMMNGTCHDDSLDSDCKLTCENGRCILNEKGLRCHWCWPSYSGGRCEVNH 4253
QY 471 T--Q-----T-G-----AL-----V-----H--S-----479
Db 4254 SNYQNGGTCTPSTLGRPTCICALGFTGPNCKAVKCEDSCHNGSCVVTAGNOPYCHCQA 4313
QY 480 Y--R-----G-----T-----G-----GG-----485
Db 4314 DYTGDRCQYVCHHYCVNSSECTIGNDGSVEVCVPTRYEGPKCIDKVCRCGHGCIINK 4373
QY 486 --IF-----E-----VC---W-----491
Db 4374 DNEDIFCNCTNGKIASSCOLCDGYNGGTGCTQDPETSPVCVCSTNWSGTQCEPAPKS 4433
QY 492 -----491
Db 4434 SKSEHISTRISAIIVPLVLVTLVTLVGLVCKRKRRTKTRROPINGINVEIGNP 4493
QY 492 --NA-----AG--D--K--VG---A-----SA-----SD-----G 505
Db 4494 SYNMYEVDHDSGGLLEPSFMDPVKSRVIGGSSAPKLPHTAPPIYLNLDLKGPLITFG 4553
QY 506 --S--V-----C-----VLDURK 514
Db 4554 PTNYSNPVYAKLYMDGQNCNRSLSV-DERK 4583

```

## RESULT 14

```

Q8NJX1
ID Q8NJX1 PRELIMINARY; PRT: 20925 AA.
AC Q8NJX1;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Nonribosomal peptide synthetase.
GN Name=pepI;
OS Trichoderma virens.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocrea; Hypocrea.
OX NCBI_TaxID=29875;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=22037864; PubMed=11909873;
RA Wiest A., Grzegorski D., Xu B.W., Goulard C., Rebuffat S.,
RA Ebbole D.J., Bodo B., Kenerley C.;
RT "Identification of peptaibols from Trichoderma virens and cloning of a
RT peptaibol synthetase."
RL J. Biol. Chem. 277:20862-20868(2002).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RA Wiest A.E., Grzegorski D., Xu B.-W., Zhou W., Ebbole D.J.,
RA Kenerley C.;
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
CC family.
CC EMBL; AF469045; AAM78457.1; -.
DR HSSP; O30409; 1DNY.
DR GO; GO:0048037; F.cofactor binding; IEA.
DR GO; GO:0016740; P:transferase activity; IEA.
DR GO; GO:0006633; P:fatty acid biosynthesis; IEA.
DR GO; GO:0008152; P:metabolism; IEA.
DR InterPro; IPR010071; AA adenyl_dom.
DR InterPro; IPR009081; ACP like.
DR InterPro; IPR012277; Ac:transferase.
DR InterPro; IPR008873; AMP-bind.
DR InterPro; IPR01242; Condensatn.
DR InterPro; IPR00794; Ketoacyl_synth.
DR InterPro; IPR006162; P:antn S.
DR InterPro; IPR010080; Thioester redct.
DR Pfam; PF00698; Acyl_transf_1; 1.
DR Pfam; PF00501; AMP-binding; 18.
DR Pfam; PF00668; Condensation; 27.
DR Pfam; PF0109; ketoacyl-synt; 1.
DR Pfam; PF00550; PP-binding; 19.
DR PRINTS; PR00154; AMPBINDING.
DR TIGRFAMs; TIGR01733; AA-adenyl-dom; 17.
DR TIGRFAMs; TIGR01746; Thioester-redct; 1.
DR PROSITE; PS00075; ACP DOMAIN; 19.
DR PROSITE; PS00455; AMP BINDING; 18.
DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN 11.
SQ SEQUENCE 20925 AA; 2320297 MW; C6603A269D56FE11 CRC64;

```

Query Match 71.3%; Score 2607.7; DB 2; Length 20925;  
 Best Local Similarity 8.1%; Pred. No. 2.2e-40;  
 Matches 476; Conservative 32; Mismatches 3; Indels 5343; Gaps 395;

```

QY 1 M-----SI-----SS-D--EV-----N-----9
Db 6914 MIVPSSVPCBGLTSIMVFETIELLEQLSSRYDAFQELPKAEPNAAVVLFTSGTGK 6973
QY 10 -----F--L--VY-----R--Y-----15
Db 6974 PKGLMEHSAFATSLGCGIYNLSPASRVQFSNIFDGLGEITFTLSFGTVCPSE 7033
QY 16 --LQ-----E-S-----19
Db 7034 DERLQKAPSFREARVNTAMLTPTSFVTRTAPQVPSLRLLVLGGEPSKDLLETWCGRLR 7093
QY 20 -----G-----F--S-H-----SA-----25
Db 7094 LVNGYGPABACNAYATTHDFKPTDSPHTIGRGFNSACWIVDPTDYNKLTPIGCIIGLIQG 7153
QY 26 -----F--T-----F-----28
Db 7154 NALARGYINDARTKNSFITNVDCLPKSIISGPHRPLYTGDLVRYTPDGOLEYLGRKDTQ 7213
QY 29 -----G-I--KS-----H-----I--S-Q--S-NI---39
Db 7214 VKLGRQELGEIEYHVKKSLANIEHVAVDVAHREGTDTLIAFVSPKMKWATTSNILL 7273
QY 40 N-----G-----A-----42
Db 7274 NDDLRLVALATIMEHLKMSLPGYMVPSTILPVRMPFITSMKVDKRLTAMAAALSLEIT 7333
QY 43 --LV-----P--P-----46
Db 7334 SPSLVKRDYIASTPMKLNLANWAOVLKISSDICKNDKSFQIGGDSSTSHLSLAQ 7393
QY 47 -----A-----AL-----I-----S 51

```

Db 7394 SGISLTVASIFADSKLSTMAISAGDGKIEPIROVEPFSPMAPTRDLDTLLIEDVRACKGLSS 7453  
QY 52 --I--I--OKGL--Q-Y--V--EA--E-- 63  
Db 7454 TAIVEDIYPTRFQEGMALAVKQPGSHIAKQVYRLSQNVVDIARFKAAWMTWELCSNLR 7513  
QY 64 --V--SI--N--E-- 68  
Db 7514 TRIVLSDASIQVVLKNVRWEDTSNMTLSYLETNQNSHMYGSALSRRHALIEEDGNH 7573  
QY 69 --DG--TLF--D--G-- 75  
Db 7574 YFVMSIHAFVDCGWTTRILNLTLSAYMGDEITPLNPTFRINVTYRSLDADAIAKIYWGEO 7633  
QY 76 --R--P--IE--S-I--S-L-- 83  
Db 7634 LLDKRALFPTISNGVPRNKSNTVRFEKAIELFRFKQTSITWASILRATWSIVLAQHCD 7693  
QY 84 --I--D-- 85  
Db 7694 TNDVTFGTALSGRQAPVPGITEKAGPVVATVPVRVRLDKLASVKFLQSIQNOASEMIPF 7753  
QY 86 --A--V--M--P--D-- 90  
Db 7754 EQFLQNIKLVNKEACDFSSLLVIQPMKLLDGNRKPILHMSATTEDSEDHMQNY 7813  
QY 91 --VVQ--T-- 94  
Db 7814 FSYPLVIQGHVSDSVNLVLIYDADILPEOQLLAHAHQFHVAAQLVAEDSGTMPLGNVS 7873  
QY 95 --R--Q--A--Y--R 100  
Db 7874 VSGSDIEYALRONNEIPRLIESCFHTVIEQIIIRPESAVHVGWGNFTYQGDDQANR 7933  
QY 101 --D--K--LA--Q--Q--Q-- 107  
Db 7934 LANHLIADYEIKNDLHVCFKSSFWFVAILAINKAGGAWPLDPSPHTQRHQIIVDQT 7993  
QY 108 -A--A--AA-- 111  
Db 7994 KARLALVSPSNI STCIDLVEHVVEVSPVTDIELSKTESHRGPDREISPSNAAYVLTSG 8053  
QY 112 --A--A--AA-- 115  
Db 8054 STGTPKGLVMEHRSVCTSLTATKRLKIRPSARTLOFAAHVFDCAVGEIISLFTGGCLF 8113  
QY 116 --A--A-- 117  
Db 8114 VPSDHRMNLPEFIRQKINYNMSTPSTFIRTLSPMDVPISLELVLVGAEVTRDIMDTWF 8173  
QY 118 --A--S--Q-- 120  
Db 8174 GKVLINGWGPACTCVVSTTHEWQSIDESPLTIGRSGVGGCWI VDPEDOKIAPVGAUGE 8233  
QY 121 --Q--GS--A--K--N-- 126  
Db 8234 IVIQPTILREYLADPVRTSGSTVYNLPNAPNRDKYWNRFYKSGDLGLYNADGTIQFA 8293  
QY 127 -- 126  
Db 8294 SRKDTIRKIRGLRVELSEVHHVKTALSRPSPGCCRCIRSTRWCQVWPHTFALAMKQGL 8353  
QY 127 --G--E-N--TA-- 131  
Db 8354 LVLAEGFDECDPFIEIEMDNGHQRUTAVISQINLSLPRYMWPTFFIPCOYMPSTSRKLD 8413  
QY 132 N--G--E--EN-G-- 137  
Db 8414 NYLRRRTAALSQEQLVTSYLGKKRSPGTVMERILQAIWRSILCIPSENIGLDDSLG 8473  
QY 138 --A-H-- 139  
Db 8474 LGGDSITAIHLVMTSREGGISLTAQIDFDDPRLFAVARAKKAGVEGGSLEIPFPFSLLS 8533

QY 140 ---T---IA---N--- 143  
Db 8534 KRSQELVITDGVRAQCDLSSETIMDAYPCTALQEGLMALSVKQPGSYIAKYVYRIFSNVD 8593  
QY 144 ---N---HTD---M---M- 149  
Db 8594 IRRFKAAMEQTIALCDNLRTIRVLVDGRATQIVLBEPIAWTVSH-DLNTAVSEMRSVKMT 8652  
QY 150 ---E---V-DG--- 153  
Db 8653 YGSPLSQFACIQBNNGDLFFLWSIIHAFDGMTVPILMNTLHAFYRGIVTPPASVARFV 8712  
QY 154 ---D-V---EI---P--N---KA-V 163  
Db 8713 KFAMSVDNEDTVKYWKDQLDNAKKATFFPAASSQNSVSULEKSITCSQTAGSSITKATI 8772  
QY 164 ---VL---R--- 166  
Db 8773 LRAAWALVLSRYSESDDVTFGATISGRQAPVDGITNMAGPAVATVPVRVICIDKKQTAAAF 8832  
QY 167 ---G---H-E---S---E---V--- 172  
Db 8833 LPSVQSQALDMISYEQYGLQNTISKISSHAMEACDFGSLIIIOPIEOLAYRGDDALLVAAE 8892  
QY 173 ---F---I-C--- 175  
Db 8893 SHITNETSALQNYFSYPLVIOCHLYDNTINMLIYDSQALSSESRMEALSHHPDNVQELA 8952  
QY 176 ---A-W---N-P--VS---D--L---LA-SG--S- 188  
Db 8953 SKONLLGSLISLAGRWDFEQALKLNSDSPDAVERCFHEMLDETALVRGDSLAI SGWDKSF 9012  
QY 189 ---GD---ST--- 192  
Db 9013 TYREMAETNRIAHLVTDYGVKVGDIHVCPEKSAWFIATLAINKAGAAMSTLDPSPH 9072  
QY 193 ---A---R---I---W---N--- 197  
Db 9073 TERYQKIIQTGSOIALSSPVNSNRCASLLPNVIELTSELDSKLAQNWEVSTSRPAVNT 9132  
QY 198 ---L--S---E---N---ST-S--- 204  
Db 9133 PRDAAYVLTSGTGVKPGVIEHGHSLCTSOISLKALDFNEFRVLQFSSSYFDILFE 9192  
QY 205 -GST- -O--LV--LR-H-- 213  
Db 9193 IGSTFLTGACLFVPSWDEQNNELVEYIRKHQLTFMVLTPTLARTIRPEDVPSVDMLVVAG 9252  
QY 214 ---C---I-R---E- 217  
Db 9253 EAPTRDILDWFGKRLANGWPTCECVIACLHQTSDVDESPKVIQRPIGGSCWIVDPED 9312  
QY 218 ---G-G--Q--D---VP--S--- 224  
Db 9313 ATCMAPLGTGVEIIOGRNLLREYLSDPVKRTATATITGLPQWPKRDSIHWDRFYLTGDL 9372  
QY 225 ---N--KD-- 227  
Db 9373 GFINEAGNLEYCTRKDTQVKIRQORLEIGEIHIIQANLESVROQAVDVTKSDAGSTLVA 9432  
QY 228 ---VT---S-LD-- 232  
Db 9433 FVSFSDATDQTVTDTRVIANGPFTSLDGDFOATISRLIGILSTLMPRYMVPSAFIPCAK 9492  
QY 233 ---W---N--- 234  
Db 9493 GEFQHTGGRYKRLRELAASLSQEDMAAYSLANEEKVAPQTMESRIQKIWAQVNLISIDS 9552  
QY 235 ---S-E-G--T---LIL- 242  
Db 9553 IGRDSDFLQIGGDSVVSIRLVSVARDANIELTVGDI FDDPRLLA VAAKATEDWRDGLMT 9612

QY 243 ---GS---Y---D---G--- 247  
Db 9613 QIBPFVLLDGSVKNVILSONIREQYSLPADKELEDAYPCTKLOGLMALAVKQPGSYIAK 9672  
QY 248 FA-RI---W---T---KD---G--- 256  
Db 9673 FLIRISNVVSHLRASWEETVLLPNLRTIRIFTNVNSVOAVFKDDVDVSWQSTSGTSLRS 9732  
QY 257 ---N--LA--S---TL---GOH---K---G--- 267  
Db 9733 YMCSTENFQAYGSPALARYALIEDGG-HTYFILLSMHHAHVDGWMRVFMAVPHSIFHNRT 9791  
QY 268 ---P---I---P--A---L---K--- 273  
Db 9792 HTVVEPYRVFKVYTLAIDSDAAADYMRSQFEGARQSIFPPTLSSVHEKKNSTLTLEKVI 9851  
QY 274 ---W---N--- 274  
Db 9852 DLPSTKSSITTATMLRAAWAIIARYCDSDDITFGATISGRQAPVPEGLADMTGPAIATV 9911  
QY 275 ---N---K--- 276  
Db 9912 PVRWINRNQIISDYLOAIQOANDMIPFEQFGLQNIKSLSDAKACDFSSLLVTPIQ 9971  
QY 277 ---K---G---NFI--- 281  
Db 9972 SLSYVDENADAIFVOAEVEKEIGTVQNYFSYPLVIOGHVHENFINLVLYDSNVLSNQ 10031  
QY 282 ---LS---A---G---VD---K--- 288  
Db 10032 ITALSHQFESVMKQASQSDMKLGSVMSASDWDLEHSMRQNRDIPDIVDSICHELIIQYQA 10091  
QY 289 -T---T---T---I-- 291  
Db 10092 VTQDPAPIVSWDRDFTYKQLEASNRHLALLVKNVNVKDDLLIPVFFEKNAVYFAITA 10151  
QY 292 I---W---D--- 294  
Db 10152 INKAGAAWVPLDPSHPVLRQLRQILSQTGTTTALSSANAVLCSTLVKRVVNAELDKL 10211  
QY 295 ---AH---TG---E---AK--- 301  
Db 10212 LATESAHGPVVDVSSRNAAYVLFTSGSTGIPKGLIMEHGSVCTSQVATAKRLGLNSKVR 10271  
QY 302 --Q---Q--- 303  
Db 10272 ILQFAAFVPLDSIGVPLISGACICVPSEHIRKNSIANFINRQGITWTYLTPTSPVRTI 10331  
QY 304 ---F---P---F---H---S--- 308  
Db 10332 KASEVPNVKLLLLAGEAVPRDIFATWFGKRLINGWPAETCCFSTLHEMAICGRESYVC 10391  
QY 309 ---AP--AL---DV---DW--- 316  
Db 10392 RRPVGFCWIVDPENPHRLAPTALGALGEVIIQGPITILREYLSVDVDRTEAAVIKSLPDWAPF 10451  
QY 317 --QS---N--- 319  
Db 10452 REQSSGRFNKSGDLGVNPDGTIEFSSRKDTQVKIRGLRVELGEVEHAQVALDGVHQI 10511  
QY 320 ---N-T---F---A--S--- 324  
Db 10512 AVDVFKDNGTQPRGYFSDESQIHEAHPSGPFQAIKQALRTVAIGKLIALPRY 10571  
QY 325 ---C---ST--D---M---C--- 330  
Db 10572 MIPTLFPCKMPSITSTKLDNRNLRRTLLTLQSELAMFSYRCNKRTPETPMESLIQRI 10631  
QY 331 ---IH--- 332  
Db 10632 WSELHILPLDSIGRDSFLGLOGDSITAIHLGSISSRQGISITAKDVDFDDPRLWDITASKA 10691  
QY 333 ---V---C-KL-G--- 337

Db 10692 RELDTAQQSPDISPFSFLIGETRELAAGSSVKQKLLTSTOTIEDAYPCTSKI PGSYI 10751  
QY 338 ---Q-D--R---PI--- 342  
Db 10752 AKYVYRLPHQVDIGRPKSSWEHTVALHPILRTRIIMIKDTCVOLVVKDSTWENLDSGLD 10811  
QY 343 ---K---T---F---Q--- 346  
Db 10812 DEAVNATHSYMTYGSPLSRYSIHESKSGDKYFLWTAHHSIHGWSVPVFNLTLYQAYKG 10871  
QY 347 ---G---H--- 348  
Db 10872 LELGPKAYSGFIKYTMEHEQKAAGDYWRQLONAKRASPPKTGLAAKSIWTKRTRVMS 10931  
QY 349 T---NEV--NA--IK--W---D---P---TG--- 361  
Db 10932 TSLKFSASSNEIATKASVIRAAWAVILARYCESDDVCFGATISGRQASVPGLEMAPAV 10991  
QY 362 ---N---LL--A---S---C--- 367  
Db 10992 ATPVVRVOLDNDQIEISKFLQNIQSOAHMVPYEQYGLQSIKLGNDARDACDFTSLLVIQ 11051  
QY 368 ---SDD--- 370  
Db 11052 PIQORQFVSQDDALLIPVETELEDILQSYNYNPLVLOGHIYNDYADLVLYDSTVVSEP 11111  
QY 371 -MT--- 373  
Db 11112 QMTALCHFNFNIVQQLAAABDGGKGLDISIASSFDLELAQRSNGDGPQIIDDCIHIIER 11171  
QY 374 --K---I--W---SM--- 378  
Db 11172 QAKQPNRPAIDAWDGKFTYSELDRTANRLAHLHVHDYAVKVGDIHVHCFEKSXWFFVAI 11231  
QY 379 ---KO---D---N---CV--- 384  
Db 11232 LAVNKAGAAWAPPDPAHPORLKAQVASTGAKALASTANTRLCEQVWDCVVEVSTLKD 11291  
QY 385 ---H--- 386  
Db 11292 NLWTVYDNIKGPDIINVTMPDAAYILFTSGSTGVPKPGIIVMQHGAICTNQALSGWLGFDH 11351  
QY 387 ---LQ---Q---H---N--- 391  
Db 11352 TVRMLQFSSFPDVSVEIVQALMNGACVCPSEHMRNLNSLDSFVRDFNVWYLTPTSFT 11411  
QY 392 ---K---E---I---Y-TI-KW--- 399  
Db 11412 RTLKPKDPFSLKLLLAGEPFTQDVLDTWFGLPNTRFINAWGPAETCVVNTLYEWQSNTE 11471  
QY 400 SP---TG---P---G--- 405  
Db 11472 SPLKGRAYGAYVWVDVENPQRLAFTGCLGEIIVQGPPLKEYLADPEKTAATVTELP 11531  
QY 406 ---T-N--- 407  
Db 11532 EWAPRQSTWNRFTYRTGDLGFDHGMHLFPASRKDTQVKIRGLRVELGEVEHRIQGLSD 11591  
QY 408 ---N---P--N---A---NL--MLA-- 416  
Db 11592 GVRQAVDVFKTEKGANLVAYLCFTSDTKTPQNTDPEGKDVFPASIDTELOINLGRMLTE 11651  
QY 417 --SAS---F---D---S--- 422  
Db 11652 LNS-SLPTMYIPTMFPICAYMPFITSSKLDVRKLRLTAELSDQQLAYSLVNEKQAPE 11710  
QY 423 T---VR---LW--- 427  
Db 11711 TEMEIRLOKLWAEVLDLPEASIGRHDNFMIRIGDSDIAAIRLVSMARDAGISLTVNDIFDD 11770  
QY 428 --- 427

Db 11771 ARLISVATKAIDNEDSHLMAPTEPPFSLTSGIEDPALPSELSGQGFIEDAYPCSKLQEG 11830  
Qy 428 -----DVD-----R----- 431  
Db 11831 LMAIAIKQPGSYIAKYIKLSEHVDVDFKFAWESTVEIASALRTRIOTGGCSIQVIN 11890  
Qy 432 G-I-----C-----I-----H----- 436  
Db 11891 GDSWDEADGSLQSLAQSTLEMGYGLRCLYRLIRRDNDTYFLNMHHAVIDGLSTQN 11950  
Qy 437 ---TL-----T---K---H----- 441  
Db 11951 ILGTLFNIYSGVDVLPPLPPYRFTQVMQLQEDTTANYWKNQLHNAQRTTFPPNDVARDK 12010  
Qy 442 -----Q---E-P-----V---Y---S---VAF---S----- 451  
Db 12011 PAATQMLQSSIELPHGLDNSTVTIATVIRAAWAIVLARYCSDDVTRGTTISGRQAPIPE 12070  
Qy 452 -----P-----D---G---R-----Y----- 456  
Db 12071 IMGWGTIATVPVRVLINRQQLISDFLEGVQROAVEMIAFEQVGLQNIAGLGDARDAC 12130  
Qy 457 -----L---AS-G-----SF- 462  
Db 12131 DFSSLLVQPIQHLAKGTKDSSILVDASIGEDNAVEAIQNVFSYPLVIAQHLVDHISFI 12190  
Qy 463 -----S---Y-----D-----K--- 464  
Db 12191 LIYDSSIMLPQMVALSYHLQVMMQLTNTTPATLETVSVSSYDVERALAINAEIPEVI 12250  
Qy 465 ---C-----VH-----I----- 468  
Db 12251 DTCHEMFENRVHLNPLAPAAWADGNYTYAELNSAANKLAHYLIKSYSKLANDLVHVCF 12310  
Qy 469 ---W-----N-----T-----QT-----G-----ALV-- 477  
Db 12311 DKSAWIYVSIIAINKAGATWPLDPHPHQRLRSIVNQTKTLALASPGNITLCSALVDN 12370  
Qy 478 -----H----- 478  
Db 12371 VVEVSSALDMLPAQDGLSPKVSRSRTAAIYLTSGSTGTPKGVVIOHRSCLTNNAI 12430  
Qy 479 -----S---Y-----R---G----- 482  
Db 12431 AKRVYHLDVRILOFSAYVDFPSLEIIMSLLQACVCPSEHIRMNGIYDFIRDMNINW 12490  
Qy 483 -----T-----GG-----I-- 486  
Db 12491 LYLTPSLRTINPIDVPNVELVAFGEAI PRDVFTWGRVRFNGWGTETTIVGSIHE 12550  
Qy 487 FE-----V---CN-----N-----A 493  
Db 12551 FESVDESPSTIGHVPGFCWIVDPNPNQLLAPTGLTGLIEIVIQGFTLLHLYLDNPKTQEA 12610  
Qy 494 -----A-----GD-----K-----V---G----- 499  
Db 12611 ILYDLPWARPDPENWGRFYKTDGLGFYNANGKIEPSSRKDTQVKIRGLRVELGEIEYQ 12670  
Qy 500 ---AS-----A-----SD-GS-V---C-----V-----LD----- 511  
Db 12671 VQASVEIRQIADVKTNGSNLVAYLCFNFDEIRQLHNDVNGPFPSPLDNLQETLAGA 12730  
Qy 512 -----L-RK 514  
Db 12731 IGKLSVTLPRYMTPTFYPCSYMPSITSGKLDK 12764

RESULT 15

Q6XA09

ID Q6XA09 PRELIMINARY; PRT: 7191 AA.

AC Q6XA09;

DT 05-JUL-2004 (TrEMBLrel. 27, Created)

DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nonribosomal peptide synthase.  
GN Name:NRPS1;  
OS Alternaria brassicae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;  
OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.  
OX NCBI\_TaxID=29911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14727058;  
RA Guillemette T., Sella A., Simoneau P.;  
RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria  
RT brassicae and flanking genomic sequences.";  
RL Curr. Genet. 45:214-224 (2004).  
CC -|- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme  
CC family.  
DR EMBL; AY246697; AAP78735.1; -;  
DR InterPro; IPR010071; AA\_adenyl\_dom.  
DR InterPro; IPR009081; ACP\_like.  
DR InterPro; IPR000873; AMP\_bind.  
DR InterPro; IPR001242; Condensatn.  
DR InterPro; IPR000977; DNA\_ligase.  
DR InterPro; IPR006162; Ppantne S.  
DR InterPro; IPR006163; Pp\_bind.  
DR Pfam; PF00501; AMP-binding; 4.  
DR Pfam; PF00668; Condensation; 9.  
DR Pfam; PF00550; PP-binding; 6.  
DR PRINTS; PR00154; AMPBINDING.  
DR TIGRFAMs; TIGR01733; AA-adenyl-dom; 4.  
DR PROSITE; PS00075; ACP\_DOMAIN; 7.  
DR PROSITE; PS00455; AMP\_BINDING; 2.  
DR PROSITE; PS00697; DNA\_LIGASE\_A1; UNKNOWN 1.  
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN 3.  
DR PROSITE; PS00012; PHOSPHOPANTETHEINE; UNKNOWN 3.  
SQ SEQUENCE 7191 AA; 793216 MW; F908905155F2394C CRC64;

Query Match 71.2%; Score 2605.9; DB 2; Length 7191;  
Best Local Similarity 8.6%; Pred. No. 3.5e-41;  
Matches 452; Conservative 48; Mismatches 10; Indels 4751; Gaps 366;  
Qy 1 M-----S-----IS-----S---D-----E----- 7  
Db 1521 MAFRLAGSGRAPAGHAETVTLARLWASVLKLDAGAVGAEDSPFRFGDSISAMKLVTAAR 1580  
Qy 8 ---V-N---F---L---V---Y---R----- 14  
Db 1581 KDGVLNVASVFAQPKLLAATAVMLPSDDAAKPEADTLPMLLPASSRQAIVALAASE 1640  
Qy 15 -----Y---LQE----- 18  
Db 1641 CDVFPDCIEDMYPCSKLQGLVMTNKPDTGVVQPIYRLPADMDISRFEAKVVAEE 1700  
Qy 19 -----S---GF-----S---H---S-----A-ET----- 27  
Db 1701 ASLRTRIVYSEEHGLQVVRREDIKWQSLPDIQHINETTRRLPAKNGAPLATFTLVGENT 1760  
Qy 28 ---F-----G-----IK---S-H-----IS-----OS-- 37  
Db 1761 DSMFFVTAHAYVDCWSWTALPRKVEANVRSGVHSPATVPYSRPFVKYISSLDQOSDA 1820  
Qy 38 -----NI-----NGAL---V---P-----P-----AA---L-- 49  
Db 1821 FWLTQLDNVTAAQFPQLPSPDHRVEANGQLLHTVLLTRNPGLEVTFVPSMIRAAWGILLAT 1880  
Qy 50 -----S---I-SII-----OK----- 55  
Db 1881 YGSGDDVINGETNSGREASVPGIESIIGTFTTAPVRLNRLNLSLTVDHYLKETQKOSLS 1940  
Qy 56 -----GLO-----Y-VE-----AE-----VS---I----- 66  
Db 1941 LPYQFAGLQHIKSLSPETAICDFQSLGLEAGDDFDDAESSLWMVYNTIGTDFSYAF 2000  
Qy 67 -----N-----E-----D-----GT----- 71

Db 2001 VFCKVNAVQVEALLFDDRRVQWTLAQRWVOQDFILTLQNLGTDNLRSLDLDLINFPA 2060  
QY 72 -----L-----F-----D----- 74  
Db 2061 DRKTISSWNSKVPPIIPRCHSVIAEDQALRPTACAIADWDTGFMYSYRELDERSALAH 2120  
QY 75 -----G-RP-----I----- 78  
Db 2121 QILRLGVRPKQFVPLCFDKSGMTVAIAVLKAGAAFPVPLDPEAPVLRLEIVSDIDADL 2180  
QY 79 -----E-----S-----L-SL----- 83  
Db 2181 LLCAPQVRELCSIPCSMVVDRQATETVAGRLPSFVSHSDPAYAFVTSSTGSKPKGA 2240  
QY 84 -----I-----DA-----V-----M-----PD----- 90  
Db 2241 VVHHTHWTSSTAFAGWKISTASRVLOFASYTFDACLIEVFSTLMQGGTVCVPDGSR 2300  
QY 91 -----VV-----Q----- 93  
Db 2301 NDLVGVRNFVNWAALTSPVSRMIVPSEIPOLETPLVGEAMSQDQDLVTWADKVNLGNG 2360  
QY 94 -----T-----R----- 95  
Db 2361 YGTECAAVATSNIMTPHMRPNNLGRAVTARGWIVSRNNHHTLAPVGAIGELLEGAVG 2420  
QY 96 -----Q-----QA-----Y-R-D----- 101  
Db 2421 AGYLNPEKTAQVFCQARWCUGLMGDDISAPVRIYKTDGLVKYNEGDTWLYLGRKDLQ 2480  
QY 102 -----KL-----A-----Q----- 106  
Db 2481 KYRGORLELSEVEHKLDDHMQVQALASVPTTGPCAKRLVAIVSLQKHGDHDDTDKQLR 2540  
QY 107 -----Q-A-----A-----A----- 113  
Db 2541 LLFQENASLNIATIRDCGLCERLPAYMIPSLWIAVERFPLMPGSKMDRRCAIQWLEQMDQA 2600  
QY 114 -----A-----A-----A-----AA----- 118  
Db 2601 TYRLISAMGTDDAEVNGSPIEKLRRAIFAKVLNLSVADVRLNQSFHLGGDSIAAMQVS 2660  
QY 119 SQ-----Q----- 122  
Db 2661 SQCRAGFPISVDIIRSKSISAWASVLDSSQSVTTEAKDYDLFPDLSPIQKVFEDAV 2720  
QY 123 -----SA-----KN-G-----E----- 128  
Db 2721 GOKHNYFNQTELPRLSRNIETELRSALTVLAKTHPMLRARFSEKNEAGAWKORIEKDVSS 2780  
QY 129 -----NTAN-----G-----E----- 134  
Db 2781 SFRLRHHVQAGNDANLRPIIDHSQATLDVAKGFTFAIELFDVDDFTFSQAIALVAHHLII 2840  
QY 135 -----EN-----G-A-----H-----T-----IA----- 142  
Db 2841 DVVSGILLLEDLQGLQLOPQSPMPYHMLHEQSLQATQESARRVFPVPGDIAPGDLDY 2900  
QY 143 -----N-N-----H-T-D-WM-----E-----V-D----- 152  
Db 2901 WAMEGRPNVNGDVVEEDHLSTRDTMLLLQAQDALATETIDILIAALFESFRKVPFDRST 2960  
QY 153 -----G-----D-----V-----E-----I-----P----- 158  
Db 2961 IAHNHGHGRETFNRRQDLRSRTVGWFTVTPHLPVPLNEATDMISTIRWRDRFRNRTPD 3020  
QY 159 -----P-----NK----- 161  
Db 3021 KGRPYFAYRNLTGQTRFASHWPAEAVFNVYGRLOSQDNKQGLFTALSDVDSREVGEDV 3080  
QY 162 -----V-----AV----- 164  
Db 3081 PRLALFDITAAVSQGAIKLSFGWRNMRQKEIRAWVGKCRQTLVDVAEBELLQARQERSV 3140

QY 165 -----L-----R-----H-E----- 169  
Db 3141 GNFKYLPLLYNGISRLSAILPAGINLVNVEDIYPASPMQOGLLHTQSRHPELYTHVTSQ 3200  
QY 170 -----S-----E-----VFI----- 174  
Db 3201 VQSGADGNPIDPRRLAEAWQVWVHRHQALRTIFIDSLAKDGSKDQIVLKEKPGRVQILADC 3260  
QY 175 -----C-A-----W-----N----- 178  
Db 3261 DDSQVANLLRHOSIDCREALPPHRMSICKTKTRVWFVKLELSHVINDGTSVSNLLADLA 3320  
QY 179 -----P-----V-----S-DL-L-A-----SG-----S 188  
Db 3321 RAYARKLTRADAGPLYSDIYGWMLSRSSDADLAYWKAHLSGIEPCLFPVVLNDGIPSPSPES 3380  
QY 189 G-D-----STARI-----W-----N----- 197  
Db 3381 GSVDELGSTRVQDFCKQNGVTLSNVQLTALTLHYVYVGTFDVSFGLIASGRDIPVTN 3440  
QY 198 LSE-----N-----S-T-----SGST-----OL----- 209  
Db 3441 INEAVGCFVNMVSRSLFSFDETTIAQLLEALOTGSTEALSHQGCSLADIQHALQLPSLFN 3500  
QY 210 -----V-----L-----R----- 212  
Db 3501 TAFTPORRSLSDPEDTALIYEDMEADAGEYIVTVNADVTQDQITVDFGYSKDRILPSQ 3560  
QY 213 -----H-----C-----I----- 215  
Db 3561 AQNMAETFKILDSIVVCSASELTIGKLDILTESSHQIMEMNPQPPPIRRCLHDVHID 3620  
QY 216 -----R-----EG-G-----QD----- 221  
Db 3621 QALTRPTTKAVEGMDGTFTYQDFDKITNQLAVHLOSIGVTTTETFPVILFEKSYAIVSM 3680  
QY 222 ----- 221  
Db 3681 IAIMKAGAYVPLDPKHPQTRBELIEDVQASVVLCSRGYHTIASEVAKTAVIVDQRSIR 3740  
QY 222 -----VP-S-----N-----K----- 226  
Db 3741 KLGVPISSPRTCAITPDNAAYCLFTSGTTGPKGTIIIPHOAFCTSAATFRRMNINATSR 3800  
QY 227 -----D-----VTS-L-----D-----W----- 233  
Db 3801 TFQFASVTFDASCIIEILSALTVGATVCVPTEDDRMNAAGAIRKLRVNMSSLTPSVLGTI 3860  
QY 234 -----N----- 235  
Db 3861 EBERVPLKTLVSGGEALSGPILKKNWSNCTCFINAYGPTCSVVAATAYKSTLDHKLIVS 3920  
QY 236 E-GTLATGS-----Y-----D-----G-FAR-----I----- 251  
Db 3921 EFGT-IGTSGCRLMIVHPRNHDKLPVSGVGLVIEGTARGYLNDVKTAKAFINDP 3979  
QY 252 -W-----T-----K-----DG-NL-----AST-----L-GO-----H----- 265  
Db 3980 AWAKTIFSSNNTFEARMYKTGDLVRYNTDGSVNYIGRKDTQIKLNGORIELGEBFHV 4039  
QY 266 ----- 265  
Db 4040 KNFPERVQAVELVAPNSRSSAKALAVFAVQDQIDGEOVSQVQASTDLPAADLLPL 4099  
QY 266 -----K-----G-----P-IF-----ALKW----- 274  
Db 4100 SDELDMCKNTENGLAGSLPSYMIPIAIFIPVTKLPWTSAGKLDNRNLSRVQNLSETWA 4159  
QY 275 -----NKK-----G----- 278  
Db 4160 MYRLTSIANKKKPITEAEKKIHKAVCVLSLPPSSVGIDDSFVRLGDSSTSMRLVAMAH 4219

QY 279 -----N-----F-----I----- 281  
Db 4220 TEQMEISFIDIFKNPKSLDLAKTGAQISKSQAQKQWQPDLLPASLTRSDVISEVVQOC 4279  
QY 282 -----LS--AG-----V-----DKT----- 289  
Db 4280 QVSKEDLQDAYPTSSLDALLTISIKQAGAYVAQHLALPKSLDMTKFKAAWESAIOEID 4339  
QY 290 --T-II-----W-----DA-----H-----T-- 297  
Db 4340 ILTRIIQMPGIFMOAVLENDPVDWREAKSLKSAEDDASKIPPHLGHLAAYTLVTTPS 4399  
QY 298 GE-----A-----K-----Q----- 302  
Db 4400 GERFYFWTLHHALYDGSIIYLMQVQOIIYSGKGVSTTPQTSYARFVEYLSSTSUSDSVY 4459  
QY 303 -----QFP-----F-HS--A-----PA----- 311  
Db 4460 WRERLTGVNAYQFPRPSHATSSAPPNGQMFQHSKMTAHRKNTDVTTPANARAAWALILAA 4519  
QY 312 -----L-----DV----- 314  
Db 4520 YTGSDDDVFGETLAGRDVAMTGDVCGPTLTTPSRVKIDRGATVSDLLNTIATNIDR 4579  
QY 315 -----S-----N-----D-W-----Q----- 317  
Db 4580 IPIQHGLSAIKALGEDMIAACDFQNLIVIQTENEBLADSMWSVHONEEQGNFTYPLVI 4639  
QY 318 -----S-----N-----T-----F-----AS----- 324  
Db 4640 ECKWGLSKTEVLAHFDANVISLHWQRLIYQFETVLIQLOSATHVRHIAVLSDQDKQVR 4699  
QY 320 --N-----T-----F-----AS----- 324  
Db 4700 KWAYEPRLIDDTVPSSLFFKKAASQPTTAVTAFDGEFSYSGELSALASQAOLVKFGAG 4759  
QY 325 -----C-----S-TD----- 328  
Db 4760 PECLIPICVDKSRWAIVAILAILISGAYVPLSPTDPASRHLHIVETCKASIVLCSPKYT 4819  
QY 329 -----M-----C-----I-H----- 332  
Db 4820 HRFVEMGVHVSSETAIRQLPTSSISLSQRAKGNICVIFTSGSTGLPKGVVIEHXS 4879  
QY 333 -----VCKLG-----Q-----D-----R-----P----- 341  
Db 4880 VSSAAICE-GLHITPTSRVFQFCFLFVSVGETLTVLLRGATICVPSEQRTNLAA 4938  
QY 342 IKT-----F-----QG-----HTN----- 350  
Db 4939 V-TDLNANWAFLPVSASTLEGPKSVPTLETLVVGGEAMTSDVVDKWATGVNLH-NGYGP 4996  
QY 351 -E-V-----N-----A-----I-----K-----W-----D-----P-----T----- 360  
Db 4997 TEGTVFAIGNDHVSAORDPSNIGHPLKSGRAWLTNSDNPHELAPIGATABELCEGPILLAR 5056  
QY 361 G-N-----L-----L-----A-----S----- 366  
Db 5057 GYLNDPKRTSEAFAPFLKFNFSNSESRIYRTGDLVQYAADGSIQYMGKKNQIKLAG 5116  
QY 367 -----C-----SD----- 369  
Db 5117 QRIELDEIQAVVHADNNHVQVQLPKVGPCCKLTVVVSFPGTAAASAGSDHRRILSDT 5176  
QY 370 -----D-----M-----TL-----K----- 374  
Db 5177 ESLSQINRARDRLADLPVSVMVPIWIAVPRIPTLASAKLDKQVGLWLEGMDALYQRI 5236  
QY 375 -----TW-----S-MK----- 379  
Db 5237 MGAELPEDMEGPGAAALTVLIRGIWAKVILNRPVEDVPKSWLSLGGDSISAMKLLAKCRS 5296  
QY 380 -----Q-D----- 381

Db 5297 EGINLNLQILRAKSILSHLAADVKSVVILDHEQNDRPFALSPIORFYVEAGSIENSTH 5356  
QY 382 -----N-----C-----V----- 384  
Db 5357 FNOGSTLRILRYVQPAWQALNSIVECHSMLRARFSPKONNGOWQOOLVMSKVSQSYAFTA 5416  
QY 385 HD-----L--Q-----O-----H----- 390  
Db 5417 HDVSTASAGAISSQKSLDRTGPVFAVDLFLNKGHOMLFMAAHLHVIDVWSWILLG 5476  
QY 391 -----N----- 391  
Db 5477 DLEDLLGSGPVTLPRSLPFQWCKMQTSSASEITQOLTVKNOPLVVEPANFAPMGMDVR 5536  
QY 392 -----KE-----IY-T-----I-K----- 398  
Db 5537 PNVGDSERDEFVIDKETSAMAFDNHHVYKTDLVDILLAAILHSFGRVFINRKAPTLFNE 5596  
QY 399 -----W-----S-----P-----T-----G-P-- 404  
Db 5597 SHGREVWDGSLDLDSRTVGWFTTLYPVTVPIDDEDEVIHTLRQVKTTRKVASNGRPYF 5656  
QY 405 -----G-----TN-----N-----P----- 409  
Db 5657 AHRYLTEDKGERFANHPMEVLPNLYGRQBSGHSLSLSPTOVEGDDDETSVDGVKTSR 5716  
QY 410 -----LAS-----A--SPD-----L-----M-- 414  
Db 5717 MALFEISASVTEGQIQLSFMYNRYSKNOKGIRRWIAECORTLEBIEAIELAKTRDPOPTWA 5776  
QY 415 -----LAS-----A--SPD-----ST-- 423  
Db 5777 DPPLPLESYRLDRVLKTLPHAGVSPFOVEDMPCSPIDQGMILSOIKSPESYSSTT 5836  
QY 424 --VR-----L--W-----D-----V-DR-----GI--C 434  
Db 5837 FEVRKRGPFVASKVVDGKQVVARHPALRTIPIDSVCKGVFDQVWKSQSDGIVTYKC 5896  
QY 435 -----I-HT-----L-----T-----K-----H----- 441  
Db 5897 ADAELATLLESIRHSLNGKKPVLPHQAQVQTSSGKIFVKIVNVHVIDGSGSLGVIGQ 5956  
QY 442 --QE-----PVYS-----VA-----F-----S-P-- 452  
Db 5957 DLOEAYEGRLEDGPLYSAIYKYLRALPAEDAIYWKAKLRGVSPCYFFTPTRDPKPRQ 6016  
QY 453 -----D-----G-----R-----YLAS----- 459  
Db 6017 LQSLDMRFTFRDELHDLAESSNVTIANILLAAWALVRSYTNSSDVCYGYLTSGRNVPID 6076  
QY 460 -----GSF-----D-----K-----C----- 465  
Db 6077 RIENAYGAFINMLVSRIELKSATSLLEIVENQSDVFGVMPHQCSLAOPQHDLGLSGKS 6136  
QY 466 --V-----H----- 467  
Db 6137 LFNATVSIQRRSSPELTSDSGIBFEQLDGHDPSEFAITVNIDATEDDDEGVFTYWSDA 6196  
QY 468 ----- 467  
Db 6197 VTDGEAKNVSTLMAKILVQVLSNPQKOTIAELDIWVKRPAQPAVSRNLNPKRPSILRSS 6256  
QY 468 -----I-----W-----N-----N-- 470  
Db 6257 SSISRSSTGPPRTPTTFPDLPAPAPLPAETPDWSSLIRSIIVSEMVPQIVEQIVAKNLS 6316  
QY 471 -----TQ--TGAL-----VHSYR----- 481  
Db 6317 TEPTSATIOMTQMTGMLTRKTSISQGRGPSIDDAGSPRAASIHRRASIASNAENRI 6376  
QY 482 -----G-----T-----GG--I 486

```

Db      6377 QTAADMVATLGVLATEASSKVAPDFVEKKNLNLWAELELLEMEETIEQDDSFQLOGDSII 6436
QY      487 -----F-----EV-C----- 490
Db      6437 AMELVGAAREEGLSMTVADVFNKPTFADMARVVVRVAGEVIDEVMRSAGGSVAGRSAGGOS 6496
QY      491 -----W-----NA----- 493
Db      6497 RSRHLDRAPSIWSEFQDINSDPNADTKSIAPSEMPAEPGLNSRDSTMFORWQGLTTNQAR 6556
QY      494 -----A-----KV-----G----- 499
Db      6557 PAASREVSOKSLATHITIQEGVSSVNRVSMLGDPNVESVISKQVFKGGISDVFPVTFDF 6616
QY      500 ---A---S-----A----- 502
Db      6617 QSLAITGLMESKMWLNLYFLDGDGDLRLKLAQAAYRMVHAFDILRTVFPYGDRLQV 6676
QY      503 -----SD-----G----- 505
Db      6677 VLRKLQPEFIYHQTDIDTFTKDLQKORENGPKLGEAFIQFVAKQKQOTGRYRIFMRL 6736
QY      506 S-----VC---VJ-----D---L-----R---K 514
Db      6737 SHAQYDGVCMKILNALQDGYNGLPVSSAPSPGPNFVRETAK 6777

```

Search completed: January 3, 2005, 15:25:01  
Job time : 135.333 secs

**This Page Blank (uspto)**



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:53:45 ; Search time 95.3333 Seconds  
(without alignments)  
1934.128 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 3669

Sequence: 1 MGISDEVNPLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*

1: Genesecp1980s:.\*  
2: Genesecp1990s:.\*  
3: Genesecp2000s:.\*  
4: Genesecp2001s:.\*  
5: Genesecp2002s:.\*  
6: Genesecp2003as:.\*  
7: Genesecp2003bs:.\*  
8: Genesecp2004s:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	3669	100.0	514	8	Adj76263 Marker ge
2	3622	98.7	514	8	Adj75513 Marker ge
3	3613	98.5	514	4	Aab95225 Human pro
4	3612	98.4	514	6	Abc07190 Human p53
5	3350.2	91.3	577	7	Add14051 Human src
6	3350.2	91.3	577	8	Adq18019 Human sof
7	3260.6	88.9	542	5	Abp41760 Human ova
8	3250.7	88.6	459	4	Admi9860 Protein e
9	2916.9	79.5	700	4	Abb60376 Drosophil
10	2791.1	76.1	584	4	Abg21351 Novel hum
11	2716	74.0	395	5	Abp51424 Human MDD
12	2631.9	71.7	5635	5	Abp60991 Novel hum
13	2627.8	71.6	5636	7	Adj70089 Human hea
14	2627.8	71.6	5636	7	Adj83137 Human hem
15	2627.8	71.6	5636	8	Adk60205 Angiogene
16	2627.8	71.6	5636	8	Adk60506 Angiogene
17	2627.8	71.6	5636	8	Adp73129 Angiogene
18	2617.2	71.3	7064	7	Adg70546 Aspergill
19	2612.9	71.2	4599	8	Adi27168 Mouse LRP
20	2612.9	71.2	4599	8	Adi27169 Mouse LRP
21	2609	71.1	31267	6	Abg74786 Human RGS
22	2608.1	71.1	4796	4	Abbs8665 Drosophil
23	2594.4	70.7	4599	6	Abp56837 Human LRP
24	2593.4	70.7	4636	4	Aae11937 Human CGI
25	2592.2	70.7	26926	4	Aau05396 Human tit

## ALIGNMENTS

## RESULT 1

ADJ76263  
ID ADJ76263 standard; protein; 514 AA.

XX AC ADJ76263;

DT 20-MAY-2004 (first entry)

XX DE Marker gene related amino acid sequence SEQ ID NO:1515.

XX KW bronchial asthma; chronic obstructive pulmonary disease;

KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;

XX KW gene therapy; marker.

XX OS Mus musculus.

XX PN EP1394274-A2.

XX PD 03-MAR-2004.

XX PF 04-AUG-2003; 2003EP-00254857.

XX PR 06-AUG-2002; 2002JP-0029312.

XX PA 20-MAR-2003; 2003JP-00077212.

XX PI (GENO-) GENOX RES INC.

XX PT Ohtani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuwara K;

XX WPI; 2004-193155/19.

XX PT Testing for bronchial asthma or chronic obstructive pulmonary disease by comparing the expression level of a marker gene in a biological sample from a subject with the expression level of the gene in a sample from a healthy subject.

XX PS Claim 16; SEQ ID NO 1515; 241pp; English.

XX CC The present invention describes a method of testing for bronchial asthma or chronic obstructive pulmonary disease. The method comprises determining the expression level of a marker gene in a biological sample from a subject, comparing the expression level determined with the expression level of the marker gene in a biological sample from a healthy subject, and judging whether the subject has bronchial asthma or chronic obstructive pulmonary disease. The marker gene comprises: (a) a group of genes (S1) whose expression levels increase when respiratory epithelial cells are stimulated with interleukin-13; or (b) a group of genes (S2) whose expression levels decrease when respiratory epithelial cells are

Adi7316 Human sof  
Adn96832 Bugula br  
Adn96829 Bugula br  
Abb11353 Human LDL  
Aau74797 Mouse alp  
Adn1590 Human CD9  
Adn1588 Human CD9  
Adn1587 Human CD9  
Adn1589 Human CD9  
Adn1588 Human CD9  
Aar47861 Alpha 2-M  
Aar60517 Human alp  
Aam9091 Human pro  
Aau81019 Human alp  
Abp56839 Human LRP  
Abu89744 Protein d  
Add14025 Human src  
Adi27167 Human LRP  
Adi15636 Human lip  
Adn1584 Human CD9  
Abb58144 Drosophil

26 2592.2 70.7 26926 8 ADQ17316  
27 2590.2 70.6 5509 8 ADN96832  
28 2590 70.6 6940 8 ADN96829  
29 2588.1 70.5 4563 4 ABB11353  
30 2587.7 70.5 4545 5 AAU74797  
31 2586.1 70.5 4419 8 ADN1590  
32 2586.1 70.5 4419 8 ADN1588  
33 2586.1 70.5 4464 8 ADN1587  
34 2586.1 70.5 4464 8 ADN1589  
35 2586.1 70.5 4544 2 AAR47861  
36 2586.1 70.5 4544 2 AAR60517  
37 2586.1 70.5 4544 4 AAM9091  
38 2586.1 70.5 4544 5 AAU81019  
39 2586.1 70.5 4544 6 ABP56839  
40 2586.1 70.5 4544 6 ABU89744  
41 2586.1 70.5 4544 7 ADD14025  
42 2586.1 70.5 4544 8 ADI27167  
43 2586.1 70.5 4544 8 ADI15636  
44 2586.1 70.5 4544 8 ADN1584  
45 2585.5 70.5 7107 4 ABB58144

stimulated with interleukin-13. Also described: (1) a reagent (I) for testing for bronchial asthma or chronic obstructive pulmonary disease; (2) a kit for screening for a candidate compound for a therapeutic agent to treat bronchial asthma or chronic obstructive pulmonary disease; (3) an animal model for bronchial asthma or chronic obstructive pulmonary disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a method for producing an animal model for bronchial asthma or chronic obstructive pulmonary disease; (6) a therapeutic agent for bronchial asthma or chronic obstructive pulmonary disease, comprising the compound, a marker gene or an antisense nucleic acid corresponding to a portion of the marker gene, a ribozyme, a polynucleotide that suppresses the expression of the gene through an RNAi effect or an antibody recognising a protein encoded by a marker gene; and (7) a DNA chip for testing for bronchial asthma or a chronic obstructive pulmonary disease, on which a probe has been immobilised to assay a marker gene. (I) has respiratory and antiasthmatic activities, and can be used in gene therapy. The method is useful for testing for or screening for a therapeutic agent for bronchial asthma or chronic obstructive pulmonary disease. The present sequence is used in the exemplification of the present invention.

XX SQ Sequence 514 AA;

Query Match 100.0%; Score 3669; DB 8; Length 514;  
 Best Local Similarity 100.0%; Pred. No. 5.9e-56;  
 Matches 514; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSISDEVNLFVRYLOESGFSHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 60  
 DB 1 MSISDEVNLFVRYLOESGFSHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 60  
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAATNQ 120  
 DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAATNQ 120  
 QY 121 QGSAKNGENTANGEGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVFICAWNPV 180  
 DB 121 QGSAKNGENTANGEGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVFICAWNPV 180  
 QY 181 SLLVSGSGSTARIWNLSENSTSGTQVLVLRHCIRREGGQDVPSNKDVTSLDWNSGTL 240  
 DB 181 SLLVSGSGSTARIWNLSENSTSGTQVLVLRHCIRREGGQDVPSNKDVTSLDWNSGTL 240  
 QY 241 ATGSDYGFARITWKDGNLSTLQGHKGPFIKALKNKXGNFILSAGVDKTTIINDAHTGEA 300  
 DB 241 ATGSDYGFARITWKDGNLSTLQGHKGPFIKALKNKXGNFILSAGVDKTTIINDAHTGEA 300  
 QY 301 KQOPFHSAPALDVWQSNNTFASCSTDMCIHVCKLQDPRPIKTFQGHTEVNAIKWDPT 360  
 DB 301 KQOPFHSAPALDVWQSNNTFASCSTDMCIHVCKLQDPRPIKTFQGHTEVNAIKWDPT 360  
 QY 361 GNLLASCSDDMTLKIWSKQDNCVHDLOAHNKEIYTIKWSPTGTNNPNANMLLASASF 420  
 DB 361 GNLLASCSDDMTLKIWSKQDNCVHDLOAHNKEIYTIKWSPTGTNNPNANMLLASASF 420  
 QY 421 DSTVRLMDVDRGICHTLTKHQBVSVAFSPPGRYLAGSPDKCVHIIWNTQTGALVHSY 480  
 DB 421 DSTVRLMDVDRGICHTLTKHQBVSVAFSPPGRYLAGSPDKCVHIIWNTQTGALVHSY 480  
 QY 481 RTGGIPEVCWNAAGDKVGASDGSVCVLDLRK 514  
 DB 481 RTGGIPEVCWNAAGDKVGASDGSVCVLDLRK 514

## RESULT 2

ADJ75513  
 ID ADJ75513 standard; protein; 514 AA.

XX AC ADJ75513;

XX DT 20-MAY-2004 (first entry)

XX DE Marker gene related amino acid sequence SEQ ID NO:765.

XX KW bronchial asthma; chronic obstructive pulmonary disease;  
 KW respiratory epithelial cell; interleukin-13; respiratory; antiasthmatic;  
 XX gene therapy; marker.  
 XX OS Homo sapiens.  
 XX PN EPI394274-A2.  
 XX PD 03-MAR-2004.  
 XX PF 04-AUG-2003; 2003EP-00254857.  
 XX PR 06-AUG-2002; 2002JP-00229312.  
 XX PR 20-MAR-2003; 2003JP-00077212.  
 XX PA (GENO-) GENOX RES INC.  
 XX PI Ontani N, Sugita Y, Yamaya M, Kubo H, Nagai H, Izuhara K;  
 XX WPI; 2004-193155/19.  
 XX PT Testing for bronchial asthma or chronic obstructive pulmonary disease by  
 PT comparing the expression level of a marker gene in a biological sample  
 PT from a subject with the expression level of the gene in a sample from a  
 PT healthy subject.

XX Example 11; SEQ ID NO 765; 241pp; English.

XX The present invention describes a method of testing for bronchial asthma  
 CC or chronic obstructive pulmonary disease. The method comprises  
 CC determining the expression level of a marker gene in a biological sample  
 CC from a subject, comparing the expression level determined with the  
 CC expression level of the marker gene in a biological sample from a healthy  
 CC subject, and judging whether the subject has bronchial asthma or chronic  
 CC obstructive pulmonary disease. The marker gene comprises: (a) a group of  
 CC genes (S1) whose expression levels increase when respiratory epithelial  
 CC cells are stimulated with interleukin-13; or (b) a group of genes (S2)  
 CC whose expression levels decrease when respiratory epithelial cells are  
 CC stimulated with interleukin-13. Also described: (1) a reagent (I) for  
 CC testing for bronchial asthma or chronic obstructive pulmonary disease;  
 CC (2) a kit for screening for a candidate compound for a therapeutic agent  
 CC to treat bronchial asthma or chronic obstructive pulmonary disease; (3)  
 CC an animal model for bronchial asthma or chronic obstructive pulmonary  
 CC disease; (4) an inducer that induces bronchial asthma in a mouse; (5) a  
 CC method for producing an animal model for bronchial asthma or chronic  
 CC obstructive pulmonary disease; (6) a therapeutic agent for bronchial  
 CC asthma or chronic obstructive pulmonary disease, comprising the compound,  
 CC a marker gene or an antisense nucleic acid corresponding to a portion of  
 CC the marker gene, a ribozyme, a polynucleotide that suppresses the  
 CC expression of the gene through an RNAi effect or an antibody recognising  
 CC a protein encoded by a marker gene; and (7) a DNA chip for testing for  
 CC bronchial asthma or a chronic obstructive pulmonary disease, on which a  
 CC probe has been immobilised to assay a marker gene. (I) has respiratory  
 CC and antiasthmatic activities, and can be used in gene therapy. The method  
 CC is useful for testing for or screening for a therapeutic agent for  
 CC bronchial asthma or chronic obstructive pulmonary disease. The present  
 CC sequence is used in the exemplification of the present invention.

XX SQ Sequence 514 AA;

Query Match 98.7%; Score 3622; DB 8; Length 514;  
 Best Local Similarity 98.8%; Pred. No. 4.6e-55;  
 Matches 508; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSISDEVNLFVRYLOESGFSHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 60  
 DB 1 MSISDEVNLFVRYLOESGFSHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 60  
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAATNQ 120  
 DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQAYRDKLAQOHAHAHAHAATNQ 120  
 QY 121 QGSAKNGENTANGEGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVFICAWNPV 180

```
Db 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVLVRGHESEVFICAWNPV 180
Qy 181 SLLVSGSGDSTARINWLNSENSTGPTQLVLRHCIREGGQDVPSNKKDVTSLDWNSEGTLL 240
Db 181 SLLVSGSGDSTARINWLNSENSTGPTQLVLRHCIREGGQDVPSNKKDVTSLDWNSEGTLL 240
Qy 241 ATGSYDGFARITWKDGNLSTLQGHKGPFPALKWKKGNPILSAGVDKTTIWDHTGEA 300
Db 241 ATGSYDGFARITWKDGNLSTLQGHKGPFPALKWKKGNPILSAGVDKTTIWDHTGEA 300
Qy 301 KOQFPFHSAPALVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
Db 301 KOQFPFHSAPALVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
Qy 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANLMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANLMLASAF 420
Qy 421 DSTVRLWDVDRGICHTLTKHQPVSVAFSPDGRYLASGSDKCVHIWNTQTGALVHSY 480
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFSPDGRYLASGSDKCVHIWNTQTGALVHSY 480
Qy 481 RGTGGIFEVCWNAAGDKVGASASDGSVCVLDLRK 514
Db 481 RGTGGIFEVCWNAAGDKVGASASDGSVCVLDLRK 514
```

```
RESULT 3
AAB95225
ID AAB95225 standard; protein; 514 AA.
XX
AC AAB95225;
XX
DT 26-JUN-2001 (first entry)
XX
DE Human protein sequence SEQ ID NO:17352.
XX
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
OS Homo sapiens.
XX
PN EP1074617-A2.
XX
PD 07-FEB-2001.
XX
PF 28-JUL-2000; 2000EP-00116126.
XX
PR 29-JUL-1999; 99JP-00248036.
PR 27-AUG-1999; 99JP-00300253.
PR 11-JAN-2000; 2000JP-00118776.
PR 02-MAY-2000; 2000JP-00183767.
PR 09-JUN-2000; 2000JP-00241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayaishi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
DR WPI; 2001-318749/34.
```

```
XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-
PT length cDNAs defined in the specification, and for the detection and/or
PT diagnosis of the abnormality of the proteins encoded by the full-length
PT cDNAs.
XX
PS Claim 8; SEQ ID NO 17352; 2537pp + Sequence Listing; English.
XX
CC The present invention describes primer sets for synthesizing 5602 full-
CC length cDNAs defined in the specification. Where a primer set comprises:
CC (a) an oligo-dr primer and an oligonucleotide complementary to the
CC complementary strand of a polynucleotide which comprises one of the 5602
CC nucleotide sequences defined in the specification, where the
```

```
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in the
CC specification. The primer sets can be used in antisense therapy and in
CC gene therapy. The primers are useful for synthesising polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893
CC represent human amino acid sequences; and AAH13629 to AAH13632 represent
CC oligonucleotides, all of which are used in the exemplification of the
CC present invention
XX
SQ Sequence 514 AA;
```

```
Query Match 98.5%; Score 3613; DB 4; Length 514;
Best Local Similarity 98.6%; Pred. No. 6.8e-55;
Matches 507; Conservative 2; Mismatches 5; Indels 0; Gaps 0;
Qy 1 MSISSDEVNFLVRYLQESGFSHSAFTGIESHISQSNINGALVPPAALISIIKGLQYV 60
Db 1 MSISSDEVNFLVRYLQESGFSHSAFTGIESHISQSNINGALVPPAALISIIKGLQYV 60
Qy 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTQQAYRDKLAQHQAAAAAATNQ 120
Db 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTQQAYRDKLAQHQAAAAAASQ 120
Qy 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVLVRGHESEVFICAWNPV 180
Db 121 QGSAKNGENTANGEANGAHTIANNHTDMMEVDGVEIPPNKAVLVRGHESEVFICAWNPV 180
Qy 181 SLLVSGSGDSTARINWLNSENSTGPTQLVLRHCIREGGQDVPSNKKDVTSLDWNSEGTLL 240
Db 181 SLLVSGSGDSTARINWLNSENSTGPTQLVLRHCIREGGQDVPSNKKDVTSLDWNSEGTLL 240
Qy 241 ATGSYDGFARITWKDGNLSTLQGHKGPFPALKWKKGNPILSAGVDKTTIWDHTGEA 300
Db 241 ATGSYDGFARITWKDGNLSTLQGHKGPFPALKWKKGNPILSAGVDKTTIWDHTGEA 300
Qy 301 KOQFPFHSAPALVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
Db 301 KOQFPFHSAPALVDWQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360
Qy 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANLMLASAF 420
Db 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTNNPNANLMLASAF 420
Qy 421 DSTVRLWDVDRGICHTLTKHQPVSVAFSPDGRYLASGSDKCVHIWNTQTGALVHSY 480
Db 421 DSTVRLWDVDRGICHTLTKHQPVSVAFSPDGRYLASGSDKCVHIWNTQTGALVHSY 480
Qy 481 RGTGGIFEVCWNAAGDKVGASASDGSVCVLDLRK 514
Db 481 RGTGGIFEVCWNAAGDKVGASASDGSVCVLDLRK 514
```

```
RESULT 4
AAB007190
ID ABO07190 standard; protein; 514 AA.
XX
AC ABO07190;
XX
DT 13-AUG-2003 (first entry)
XX
DE Human p53 modifying protein, SEQ ID 150.
XX
KW Human; p53 modifier; cytostatic; cancer; cytostatic; antiangiogenic;
KW antiapoptotic; p53 pathway; breast cancer; colon cancer; kidney cancer;
```

lung cancer; ovarian cancer; angiogenesis; cell cycle; apoptotic disorder; cell proliferation disorder.

Homo sapiens.

WO200299122-A1.

12-DEC-2002.

03-JUN-2002; 2002WO-US017382.

05-JUN-2001; 2001US-0296076P.

10-OCT-2001; 2001US-0328605P.

15-FEB-2002; 2002US-0357253P.

(EXEL-) EXELIXIS INC.

Friedman L, Plowman GD, Belvin M, Francis-Lang H, Li D, Funke RP; WPI; 2003-156859/15.

N-PSDB; ACD13365.

Identifying modulators of the p53 pathway for use in treating apoptotic or cell proliferation disorders, comprises screening for agents that modulate activity of a human ortholog of genes that modify the p53 pathway in Drosophila.

Example 2; Page 458-459; 678pp; English.

The invention relates to identifying (M1) a candidate p53 pathway modulating agent, by contacting an assay system comprising a purified HM polypeptide (human orthologue of genes that modify the p53 pathway in Drosophila) or nucleic acid with a test agent under conditions, where but for the presence of the test agent, the system provides a reference activity, and detecting a test agent-biased activity of the assay system. Also included are modulating (M2) a p53 pathway of a cell (comprising contacting a cell defective in p53 function with a candidate modulator that specifically binds to a HM polypeptide comprising an HM amino acid sequence, where p53 function is restored), modulating (M3) a p53 pathway in a mammalian cell (comprising contacting the cell with an agent that specifically binds an HM polypeptide or nucleic acid) and diagnosing (M4) a disease in a patient (comprising: (a) obtaining a biological sample from the patient; (b) contacting the sample with a probe for HM expression; (c) comparing the results with a control; and (d) determining whether the comparison indicates a likelihood disease). (M1) is useful for identifying modulators of the p53 pathway. A probe for HM expression is useful for diagnosing breast, colon, kidney, lung and ovarian cancer, in a patient, where the cancer has greater than 25 % expression level. Modulators identified by (M1) are useful in a variety of diagnostic and therapeutic applications, where disease or disorder prognosis is related to defects in the p53 pathway, such as, angiogenesis, apoptotic or cell proliferation disorders (e.g., cancer). Another two new methods (M2 and M3) are useful for modulating the p53 pathway of a cell, thus restoring the p53 function of the cell, so that the cell undergoes normal proliferation or progression through the cell cycle. (M2) and (M3) are also useful for treating defects in the p53 pathway such as angiogenic, apoptotic or cell proliferation disorders. The present sequence represents a human p53 pathway modifying protein

Sequence 514 AA;

Query Match 98.4%; Score 3612; DB 6; Length 514;  
Best Local Similarity 98.6%; Pred. No. 7.1e-55;  
Matches 507; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

1 MSISSEVNVFLVRYLOESGFSHSAFTFGIESHSQSNGALVPPAALISIIKGLQYV 60  
1 MSISSEVNVFLVRYLOESGFSHSAFTFGIESHSQSNGALVPPAALISIIKGLQYV 60

61 EAEVSINEDCTLFDGRIEISLIDAVPDPVOTROQAYRDLKQAQAAAAAATNQ 120  
61 EAEVSINEDCTLFDGRIEISLIDAVPDPVOTROQAYRDLKQAQAAAAAATNQ 120

121 QGSAKNGENTANGEANGAHTIANNHTDMMEDVDGEVPEISNKAVVLRGHESEVFCANPV 180  
121 QGSAKNGENTANGEANGAHTIANNHTDMMEDVDGEVPEISNKAVVLRGHESEVFCANPV 180

181 SLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240  
181 SLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSGTL 240

241 ATGSYDGFARITKDGKGLASTLQHKGPFPALKWKKGNFIFLSAGVDKTTIWDHTGEA 300  
241 ATGSYDGFARITKDGKGLASTLQHKGPFPALKWKKGNFIFLSAGVDKTTIWDHTGEA 300

301 KQFPFHSAAPALVDWQSNNTFASCSTDCIHVKLGQDRPIKTFQCHTNEVNAIKWDPT 360  
301 KQFPFHSAAPALVDWQSNNTFASCSTDCIHVKLGQDRPIKTFQCHTNEVNAIKWDPT 360

361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKIYTIKWSPTGPTGNNPNANLMLASAF 420  
361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKIYTIKWSPTGPTGNNPNANLMLASAF 420

421 DSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGSPDKCVHIWNTOTGALVHSY 480  
421 DSTVRLWDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGSPDKCVHIWNTOTGALVHSY 480

481 RGTGGIFEVCAAGDKVGSASDGSVCVLDLRK 514  
481 RGTGGIFEVCAAGDKVGSASDGSVCVLDLRK 514

RESULT 5  
ADD14051  
ID ADD14051 standard; protein; 577 AA.  
XX AC ADD14051;  
XX DT 01-JAN-2004 (first entry)  
XX DE Human src biomarker polypeptide SEQ ID NO:240.  
XX KW predictor set; protein tyrosine kinase activity modulator;  
XX KW protein tyrosine kinase pathway; protein tyrosine kinase; cytostatic;  
XX KW gene therapy; drug sensitivity; genetic profile; cancer; human.  
XX OS Homo sapiens.  
XX PN WO2003062395-A2.  
XX PD 31-JUL-2003.  
XX PF 17-JAN-2003; 2003WO-US001981.  
XX PR 18-JAN-2002; 2002US-0350061P.  
XX PA (BRIM ) BRISTOL-MYERS SQUIBB CO.  
XX PI Huang F, Fairchild CR, Lee FY, Shaw P;  
XX DR WPI; 2003-636735/60.  
XX DR N-PSDB; ADD14646.  
XX PT New polynucleotides and polypeptides for predicting the activity of  
XX PT compounds that interact with protein tyrosine kinases and/or protein  
XX PT tyrosine kinase pathways.  
XX PS Claim 10; SEQ ID NO 240; 139pp; English.  
XX CC The present invention describes a predictor set comprising a plurality of  
XX CC polynucleotides or polypeptides whose expression pattern is predictive of  
XX CC the response of cells to treatment with a compound that modulates protein  
XX CC tyrosine kinase activity or members of the protein tyrosine kinase  
XX CC pathway. Also described: (1) predicting whether a compound is capable of  
XX CC modulating the activity of cells, comprising obtaining a sample of cells,  
XX CC determining whether the cells express a plurality of markers, and

CC correlating the expression of the markers to the compound's ability to  
 CC modulate the activity of the cells; (2) a plurality of cell lines for  
 CC identifying polynucleotides and polypeptides whose expression levels  
 CC correlate with compound sensitivity or resistance of cells associated  
 CC with a disease state; and (3) identifying polynucleotides and  
 CC polypeptides that predict compound sensitivity or resistance of cells  
 CC associated with a disease state, comprising subjecting the plurality of  
 CC cell lines to one or more compounds, analysing the expression pattern of  
 CC a microarray of polynucleotides or polypeptides, and selecting  
 CC polynucleotides or polypeptides that predict the sensitivity or  
 CC resistance of cells associated with a disease state by using the  
 CC expression pattern of the microarray. The polynucleotides and  
 CC polypeptides have cytoskeletal activities, and can be used in gene therapy.  
 CC The polynucleotides and polypeptides are useful in predicting the  
 CC activity of compounds that interact with protein tyrosine kinases and/or  
 CC protein tyrosine kinase pathways. These may be used in determining drug  
 CC sensitivity in patients to allow the development of individualized  
 CC genetic profiles which aid in treating diseases and disorders (e.g.  
 CC cancer) based on patient response at a molecular level. The present  
 CC sequence is used in the exemplification of the present invention.  
 XX  
 SQ Sequence 577 AA;

Query Match 91.3%; Score 3350.2; DB 7; Length 577;  
 Best Local Similarity 86.2%; Pred. No. 1.2e-49;  
 Matches 456; Conservative 36; Mismatches 19; Indels 18; Gaps 10;

QY 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISIIQKGLQYV 60  
 DB 52 MSITSDEVNFLVRYLOESGFSHSAFTFGIESHSOSNINGTLVPPAALISIIQKGLQYV 111  
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQOH---AAAAA--- 115  
 DB 112 EAEISINEDGTVFDRPIESLSLIDAVMPDVVQTRQOAFREKLAQOQAAAAAATAA 171  
 QY 116 --AATNQ-QG-S---AKNGENTANGENGAAHTIANNHTDM-MEVDGDVEIIPSNKAVLR 166  
 DB 172 ATAATTSAGVSHQNPCKNREATVNGEENRAHSV-NNHA-KPMEIDGEVEIPSSKATVLR 229  
 QY 167 GHESEVFIICANWPNVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGDVP SNK 226  
 DB 230 GHESEVFIICANWPNVSDLLASGSGDSTARIWNLSNENSGGSTQLVLRHCIREGGDVP SNK 289  
 QY 346 QGHTNEVNAIKWDPPTGNLLASCDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGP 405  
 DB 409 QGHTNEVNAIKWDPSPGMLLASCDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 468

227 DVTSLDWNSEGTLLATGSDYGFARITWKDGNLSTLQGHKGPITFALKWKNKGNFILLSAGV 286  
 290 DVTSLDWNNTGTLATGSDYGFARITWEDGNLSTLQGHKGPITFALKWKNKGNFILLSAGV 349  
 287 DKTIIWDAHTGEAKQOQFPFHSAPALDQVQNN-TFASCSTDMCIHVCKLGDGDRPIKTF 345  
 350 DKTIIWDAHTGEAKQOQFPFHSAPALDQVQ--NNTTFASCSTDMCIHVCKLGDGDRPVKTF 408  
 287 DKTIIWDAHTGEAKQOQFPFHSAPALDQVQNN-TFASCSTDMCIHVCKLGDGDRPIKTF 345  
 350 DKTIIWDAHTGEAKQOQFPFHSAPALDQVQ--NNTTFASCSTDMCIHVCKLGDGDRPVKTF 408  
 346 QGHTNEVNAIKWDPPTGNLLASCDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGP 405  
 409 QGHTNEVNAIKWDPSPGMLLASCDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 468  
 406 TNNPNNMLASAFDSTVRLMDVDRGICHTLTKHQEPVYSVAFSPDGRYLASGSDFC 465  
 469 TSNPNSNIMLASAFDSTVRLMDIERGVCTHILTKHQEPVYSVAFSPDGRYLASGSDFC 528  
 466 VHIWNTQSGALVHSYRGTTGGIPEVCWNAAGDKVGSASDGSVCVLDLRK 514  
 529 VHIWNTQSGNLVHSYRGTTGGIPEVCWNAARGDKVGSASDGSVCVLDLRK 577

RESULT 6  
 ADQ18019  
 ID ADQ18019 standard; protein; 577 AA.  
 XX  
 AC ADQ18019;  
 XX  
 DT 26-AUG-2004 (first entry)  
 XX  
 DE Human soft tissue sarcoma-upregulated protein - SEQ ID 836.

XX soft tissue sarcoma; cytostatic; gene therapy; vaccine; screening; human.  
 XX Homo sapiens.  
 XX WO2004048938-A2.  
 XX 10-JUN-2004.  
 XX 26-NOV-2003; 2003WO-US038193.  
 XX 26-NOV-2002; 2002US-0429739P.  
 XX (PROT-) PROTEIN DESIGN LABS INC.  
 XX Aziz N, Ginsburg WM, Zlotnik A;  
 XX WPI; 2004-441208/41.  
 XX Early detection of soft tissue sarcoma comprises determining expression  
 XX of a gene in a first soft tissue sample and a normal soft tissue sample  
 XX and comparing the gene expression, also useful in treating soft tissue  
 XX sarcoma.  
 XX Example 2; SEQ ID NO 836; 210pp; English.  
 XX The invention relates to a novel method for detecting soft tissue sarcoma  
 XX which comprises obtaining a first soft tissue sample from an individual  
 XX and a normal soft tissue sample from the same or different individual,  
 XX determining the expression of a gene in both samples and comparing the  
 XX expression of the gene in both soft tissue samples, where a higher level  
 XX of protein expression in the first soft tissue sample indicates the  
 XX presence of soft tissue sarcoma. The method of the invention has  
 XX cytoskeletal applications and may be useful for detecting soft tissue  
 XX sarcoma, possibly via gene therapy or vaccine production. The nucleic  
 XX acid sequences may be useful in diagnostic and screening applications.  
 XX The current sequence is that of a human soft tissue sarcoma-upregulated  
 XX protein of the invention. The current sequence is not shown within the  
 XX specification per se but was submitted in CD format by the inventor.  
 XX Sequence 577 AA;

Query Match 91.3%; Score 3350.2; DB 8; Length 577;  
 Best Local Similarity 86.2%; Pred. No. 1.2e-49;  
 Matches 456; Conservative 36; Mismatches 19; Indels 18; Gaps 10;

QY 1 MSISDEVNFLVRYLOESGFSHSAFTFGIESHSOSNINGALVPPAALISIIQKGLQYV 60  
 DB 52 MSITSDEVNFLVRYLOESGFSHSAFTFGIESHSOSNINGTLVPPAALISIIQKGLQYV 111  
 QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQOH---AAAAA--- 115  
 DB 112 EAEISINEDGTVFDRPIESLSLIDAVMPDVVQTRQOAFREKLAQOQAAAAAATAA 171  
 QY 116 --AATNQ-QG-S---AKNGENTANGENGAAHTIANNHTDM-MEVDGDVEIIPSNKAVLR 166  
 DB 172 ATAATTSAGVSHQNPCKNREATVNGEENRAHSV-NNHA-KPMEIDGEVEIPSSKATVLR 229  
 QY 167 GHESEVFIICANWPNVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGDVP SNK 226  
 DB 230 GHESEVFIICANWPNVSDLLASGSGDSTARIWNLSNENSGGSTQLVLRHCIREGGDVP SNK 289  
 QY 227 DVTSLDWNSEGTLLATGSDYGFARITWKDGNLSTLQGHKGPITFALKWKNKGNFILLSAGV 286  
 DB 290 DVTSLDWNNTGTLATGSDYGFARITWEDGNLSTLQGHKGPITFALKWKNKGNFILLSAGV 349  
 287 DKTIIWDAHTGEAKQOQFPFHSAPALDQVQNN-TFASCSTDMCIHVCKLGDGDRPIKTF 345  
 350 DKTIIWDAHTGEAKQOQFPFHSAPALDQVQ--NNTTFASCSTDMCIHVCKLGDGDRPVKTF 408  
 287 DKTIIWDAHTGEAKQOQFPFHSAPALDQVQNN-TFASCSTDMCIHVCKLGDGDRPIKTF 345  
 350 DKTIIWDAHTGEAKQOQFPFHSAPALDQVQ--NNTTFASCSTDMCIHVCKLGDGDRPVKTF 408  
 346 QGHTNEVNAIKWDPPTGNLLASCDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGP 405  
 409 QGHTNEVNAIKWDPSPGMLLASCDDMTLKIWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA 468

QY 406 TNNPNNMLASAFSTVRLWDVDRGICHTLTKHQBVPVSVAPSPGRYLASGDFK 465  
 Db 469 TNPNSNIMLASAFSTVRLWDIERGVCTHTLTTHQEPVSVAFSPDKYLASGDFK 528  
 QY 466 VHMWNTQTGALVHSYRGTTGGIFEVCMWNAAGDKVGASASDGSVCVLDLRK 514  
 Db 529 VHMWNTQSNLVHSYRGTTGGIFEVCMWNAAGDKVGASASDGSVCVLDLRK 577

RESULT 7  
 ID ABP41760 standard; protein; 542 AA.  
 XX  
 AC ABP41760;  
 XX  
 DT 22-AUG-2002 (first entry)  
 XX  
 DE Human ovarian antigen HNOKM38, SEQ ID NO:2892.  
 XX  
 KW Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;  
 KW ovarian cancer; breast cancer; tumour; reproductive system disorder;  
 KW infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;  
 KW PCOS; ovarian cyst; dysmenorrhea; endocrine disorder; infection;  
 KW inflammatory condition; immune disorder; blood disorder;  
 KW cardiovascular disorder; respiratory disorder; neurological disorder;  
 KW gastrointestinal disorder; urinary system disorder; drug screening;  
 KW gene therapy; chromosome mapping; forensic analysis;  
 KW antibody preparation; cytostatic; immunomodulatory; neuroprotective;  
 KW antiinflammatory; gynaecological; reproductive.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2002006777-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 07-JUN-2001; 2001WO-US018569.  
 XX  
 PR 07-JUN-2000; 2000US-0209467P.  
 XX  
 PA (HUMA-) HUMAN GENOME SCI INC.  
 XX  
 PI Birse CE, Rosen CA;  
 XX  
 DR WPI; 2002-147878/19.  
 DR N-PSDB; ABQ54837.  
 XX  
 PT Isolated nucleic acid molecules encoding novel ovarian polypeptides,  
 PT useful in the prevention, treatment and diagnosis of cancer (e.g. ovarian  
 PT cancer), immune disorders, cardiovascular disorders and neurological  
 PT diseases.  
 XX  
 PS Claim 11; SEQ ID NO 2892; 2922pp; English.  
 XX  
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-  
 CC ABP43228) and to cDNAs encoding them (ABQ54131-ABQ56305), and also  
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical  
 CC to the sequences of the invention. The invention additionally relates to  
 CC recombinant vectors and host cells comprising human ovarian antigen  
 CC polynucleotides, antibodies against human ovarian antigens, and the use  
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,  
 CC treating, prognosing or preventing various ovarian and/or breast-related  
 CC disorders. Such conditions include ovarian cancer and breast cancer, and  
 CC metastatic tumours of ovarian or breast origin, reproductive system  
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,  
 CC polycystic ovary syndrome, ovarian cysts and dysmenorrhea), endocrine  
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic  
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and  
 CC vaginitis), immune disorders (e.g., congenital and acquired  
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),  
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,  
 CC respiratory disorders, neurological disorders, gastrointestinal disorders

CC and urinary system disorders. Ovarian antigen polypeptides and  
 CC polynucleotides may also be used in screening for compounds which  
 CC modulate ovarian antigen expression or activity. The polynucleotides may  
 CC further be used for gene therapy, chromosome mapping, in the  
 CC identification of individuals and in forensic analysis, and the  
 CC polypeptides may be used as food additives or to prepare antibodies  
 CC useful in disease diagnosis, drug targeting and phenotyping. The present  
 CC sequence represents a human ovarian antigen of the invention. Note: The  
 CC sequence data for this patent did not form part of the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences

XX Sequence 542 AA;

Query Match 88.9%; Score 3260.6; DB 5; Length 542;  
 Best Local Similarity 80.7%; Pred. No. 4.5e-48;  
 Matches 442; Conservative 40; Mismatches 22; Indels 44; Gaps 15;

QY 5 SDEVNF---LV-----YR-----YLOESGFSHSAFTFGIESHSQSNING 41  
 Db 1 TDC--FKESKVDLQSQOKPLRTPHTLCGFCGTGFLTC-GFHSATFTXIESHSQSNING 57  
 QY 42 ALVPPAALISIIQKGLQYVEAEVSNEDCTLFDGRIEISLSLIDAVMPDVVQTRQAYRD 101  
 Db 58 TLVPPAALISIIQKGLQYVEAEISINEDGTVDGRIEISLSLIDAVMPDVVQTRQAFRE 117  
 QY 102 KLAQQH-AAAAA-----AATNQ-QG-S-----AKNGENTANGEANGHTIANNHTDM 148  
 Db 118 KLAQQASAAAAAATAATAATTTAGVSHQNFSKNREAVNGENRAHSV-NNHA-K 175  
 QY 149 -MEVDGVEIPSNKAVLRGHESEVFCANWPSVLLVSGSDSTARIWNLSENSTSGPT 207  
 Db 176 PMEIDGVEIPSKATVLRGHESEVFCANWPSVLLVSGSDSTARIWNLSENSTSGPT 235  
 QY 208 QLVLRHCIREGGQDVPSNKNVTSLDWNSBGLTATGSDYDGFARIWTKDGNLASTLQKHG 267  
 Db 236 QLVLRHCIREGGHDVPSNKNVTSLDWNTNGTLATGSDYDGFARIWTEGDNLASTLQKHG 295  
 QY 268 PIFALKKWKKNFILLSAGVDKTTIIWDAHTGBAKQOQFPFHSAPALDWDQSN-TFASCS 326  
 Db 296 PIFALKKWKKNFILLSAGVDKTTIIWDAHTGBAKQOQFPFHSAPALDWDQ-NTTTFASCS 354  
 QY 327 TDMCIHVCKLGODRPIKTFQGHTEVNAIKWDPCTNLLASCSDDMTLKWSKKDNCVVD 386  
 Db 355 TDMCIHVCKLGCDRPIKTFQGHTEVNAIKWDPCTNLLASCSDDMTLKWSKKQVCIHD 414  
 QY 387 LQAHNKEIYTIKWSPTGPTNNPNNMLASAFSTVRLWDVDRGICHTLTKHQBVPV 446  
 Db 415 LQAHNKEIYTIKWSPTGPTNPNPNSNIMLASAFSTVRLWDIERGVCTHTLTTHQEPV 474  
 QY 447 SVAFSPDGRYLASGDFKVCVHIWNTQTGALVHSYRGTTGGIFEVCMWNAAGDKVGASASDGS 506  
 Db 475 SVAFSPDGRYLASGDFKVCVHIWNTQSGNLVHSYRGTTGGIFEVCMWNAAGDKVGASASDGS 534  
 QY 507 VCVLDRK 514  
 Db 535 VCVLDRK 542

RESULT 8

ADM19860

ID ADM19860 standard; protein; 459 AA.

XX ADM19860;

XX 20-MAY-2004 (first entry)

XX Protein encoded by novel human channel/transporter gene #178.

XX immunosuppressive; antiarthritic; antirheumatic; antiproliferative;  
 KW cytosolic; cardiac; vasotropic; cerebroprotective; nootropic;  
 KW neuroprotective; antibacterial; virucide; fungicide; opthalmological;  
 KW gene therapy; channel/transporter protein; rheumatoid arthritis;

neoplasm; cardiac arrest; cerebrovascular disorder; cerebral ischemia;  
angiogenesis; nervous system disorder; Alzheimer's disease;  
ocular disorder; corneal infection; wound healing;  
epithelial cell proliferation; skin aging; sunburn; transplantation;  
chemotaxis; food additive.

Homo sapiens.

WO200154472-A2.

02-AUG-2001.

17-JAN-2001; 2001WO-US0001307.

31-JAN-2000; 2000US-0179065P.

04-FEB-2000; 2000US-0180628P.

24-FEB-2000; 2000US-0184664P.

02-MAR-2000; 2000US-0186350P.

16-MAR-2000; 2000US-0189874P.

17-MAR-2000; 2000US-0190076P.

18-APR-2000; 2000US-0198123P.

19-MAY-2000; 2000US-0205515P.

07-JUN-2000; 2000US-0209467P.

28-JUN-2000; 2000US-0214886P.

30-JUN-2000; 2000US-0215135P.

07-JUL-2000; 2000US-0216647P.

07-JUL-2000; 2000US-0216880P.

11-JUL-2000; 2000US-0217487P.

11-JUL-2000; 2000US-0217496P.

14-JUL-2000; 2000US-0218290P.

26-JUL-2000; 2000US-0220963P.

26-JUL-2000; 2000US-0220964P.

14-AUG-2000; 2000US-0224518P.

14-AUG-2000; 2000US-0224519P.

14-AUG-2000; 2000US-0225113P.

14-AUG-2000; 2000US-0225214P.

14-AUG-2000; 2000US-0225266P.

14-AUG-2000; 2000US-0225758P.

14-AUG-2000; 2000US-0225759P.

18-AUG-2000; 2000US-0226279P.

22-AUG-2000; 2000US-0226681P.

22-AUG-2000; 2000US-0226868P.

22-AUG-2000; 2000US-0227182P.

23-AUG-2000; 2000US-0227009P.

30-AUG-2000; 2000US-0228924P.

01-SEP-2000; 2000US-0229287P.

01-SEP-2000; 2000US-0229343P.

01-SEP-2000; 2000US-0229344P.

01-SEP-2000; 2000US-0229345P.

05-SEP-2000; 2000US-0229509P.

05-SEP-2000; 2000US-0229513P.

06-SEP-2000; 2000US-0230437P.

08-SEP-2000; 2000US-0230438P.

08-SEP-2000; 2000US-0231142P.

08-SEP-2000; 2000US-0231143P.

08-SEP-2000; 2000US-0231244P.

08-SEP-2000; 2000US-0231413P.

08-SEP-2000; 2000US-0231414P.

08-SEP-2000; 2000US-0232080P.

08-SEP-2000; 2000US-0232081P.

12-SEP-2000; 2000US-0231968P.

14-SEP-2000; 2000US-0232397P.

14-SEP-2000; 2000US-0232398P.

14-SEP-2000; 2000US-0232399P.

14-SEP-2000; 2000US-0233065P.  
21-SEP-2000; 2000US-0234223P.  
21-SEP-2000; 2000US-0234274P.  
25-SEP-2000; 2000US-0234997P.  
25-SEP-2000; 2000US-0234998P.  
26-SEP-2000; 2000US-0235484P.  
27-SEP-2000; 2000US-0235834P.  
27-SEP-2000; 2000US-0235836P.  
29-SEP-2000; 2000US-0236327P.  
29-SEP-2000; 2000US-0236367P.  
29-SEP-2000; 2000US-0236368P.  
29-SEP-2000; 2000US-0236369P.  
29-SEP-2000; 2000US-0236370P.  
02-OCT-2000; 2000US-0236802P.  
02-OCT-2000; 2000US-0237037P.  
02-OCT-2000; 2000US-0237038P.  
02-OCT-2000; 2000US-0237039P.  
02-OCT-2000; 2000US-0237040P.  
13-OCT-2000; 2000US-0239935P.  
13-OCT-2000; 2000US-0239937P.  
20-OCT-2000; 2000US-0240960P.  
20-OCT-2000; 2000US-0241221P.  
20-OCT-2000; 2000US-0241785P.  
20-OCT-2000; 2000US-0241786P.  
20-OCT-2000; 2000US-0241787P.  
20-OCT-2000; 2000US-0241808P.  
20-OCT-2000; 2000US-0241809P.  
20-OCT-2000; 2000US-0241826P.  
01-NOV-2000; 2000US-0244617P.  
08-NOV-2000; 2000US-0246474P.  
08-NOV-2000; 2000US-0246475P.  
08-NOV-2000; 2000US-0246476P.  
08-NOV-2000; 2000US-0246477P.  
08-NOV-2000; 2000US-0246478P.  
08-NOV-2000; 2000US-0246523P.  
08-NOV-2000; 2000US-0246524P.  
08-NOV-2000; 2000US-0246525P.  
08-NOV-2000; 2000US-0246526P.  
08-NOV-2000; 2000US-0246527P.  
08-NOV-2000; 2000US-0246528P.  
08-NOV-2000; 2000US-0246532P.  
08-NOV-2000; 2000US-0246609P.  
08-NOV-2000; 2000US-0246610P.  
08-NOV-2000; 2000US-0246611P.  
08-NOV-2000; 2000US-0246613P.  
17-NOV-2000; 2000US-0249207P.  
17-NOV-2000; 2000US-0249208P.  
17-NOV-2000; 2000US-0249209P.  
17-NOV-2000; 2000US-0249210P.  
17-NOV-2000; 2000US-0249211P.  
17-NOV-2000; 2000US-0249217P.  
17-NOV-2000; 2000US-0249218P.  
17-NOV-2000; 2000US-0249219P.  
17-NOV-2000; 2000US-0249213P.  
17-NOV-2000; 2000US-0249214P.  
17-NOV-2000; 2000US-0249215P.  
17-NOV-2000; 2000US-0249216P.  
17-NOV-2000; 2000US-0249217P.  
17-NOV-2000; 2000US-0249265P.  
17-NOV-2000; 2000US-0249297P.  
17-NOV-2000; 2000US-0249299P.  
17-NOV-2000; 2000US-0249300P.  
01-DEC-2000; 2000US-0250160P.  
01-DEC-2000; 2000US-0250391P.  
05-DEC-2000; 2000US-0251030P.  
05-DEC-2000; 2000US-0251988P.  
06-DEC-2000; 2000US-0256719P.  
08-DEC-2000; 2000US-0251479P.  
08-DEC-2000; 2000US-0251856P.  
08-DEC-2000; 2000US-0251868P.  
08-DEC-2000; 2000US-0251869P.

```
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Rosen CA, Barash SC, Ruben SM;
XX
PI WPI; 2001-476159/51.
DR N-PSDB; ADM19381.
XX
PT Isolated nucleic acid molecule encoding a channel/transporter protein is
PT used in preventing, treating or ameliorating a medical condition.
XX
PS Claim 11; SEQ ID NO 667; 809pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule encoding a
CC channel/transporter protein or sequences at least 95% identical to a
CC these. The nucleic acids and proteins encoded by them are used to
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used
CC in diagnosing a pathological condition or susceptibility to a
CC pathological condition. The antibodies to the proteins can also be used
CC in alleviating symptoms associated with the disorders and in diagnostic
CC immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays
CC (ELISA). Disorders which are diagnosed or treated include autoimmune
CC diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.
CC neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac
CC arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,
CC nervous system disorders e.g. Alzheimer's disease, infections caused by
CC bacteria, viruses and fungi and ocular disorders e.g. corneal infection.
CC The polypeptides can also be used to aid wound healing and epithelial
CC cell proliferation, to prevent skin aging due to sunburn, to maintain
CC organs before transplantation, for supporting cell culture of primary
CC tissues, to regenerate tissues and in chemotaxis. The polypeptides can
CC also be used as a food additive or preservative to increase or decrease
CC storage capabilities. This sequence corresponds to a protein of the
CC invention.
XX
SQ Sequence 459 AA;
Query Match 88.6%; Score 3250.7; DB 4; Length 459;
Best Local Similarity 91.9%; Pred. No. 3.1e-48;
Matches 452; Conservative 3; Mismatches 4; Indels 33; Gaps 3;
QY 23 HSAFTFGESHISQSNINGALVPPAALISIIQKQYVEAEVSEINEDGTLFDRGPISLS 82
DB 1 H-A-----SD-----VEAEVSEINEDGTLFDRGPISLS 27
QY 83 LIDAVMPDVVQTRQAYRDKLAQAAAAAATAATNQGSXKNGENTANGENGARTIA 142
DB 28 LIDAVMPDVVQTRQAYRDKLAQAAAAAATAATNQGSXKNGENTANGENGARTIA 87
QY 143 NNHTDMNEVDGVEIPNKAVALRGHSEVFCANPVSDDLIVSGSDSTARIWNLSNS 202
DB 88 NNHTDMNEVDGVEIPNKAVALRGHSEVFCANPVSDDLIVSGSDSTARIWNLSNS 147
QY 203 TSGPTQLVLRHCIRREGQDVPSNKNVTSLDWNSGTLATGSDGFGARITWKDGNLASTL 262
DB 148 TSGPTQLVLRHCIRREGQDVPSNKNVTSLDWNSGTLATGSDGFGARITWKDGNLASTL 207
QY 263 GQHKGPFIALKWKNKGNFILSAGVDKTTIWDATGCAKQFPFHSAPALDWDVQSNNTF 322
DB 208 GQHKGPFIALKWKNKGNFILSAGVDKTTIWDATGCAKQFPFHSAPALDWDVQSNNTF 267
QY 323 ASCSTDMCIHVCKIGDRLPKTFQHTNEVNAIKWPTGNLLASCDDMTLKIWSMKQDN 382
DB 268 ASCSTDMCIHVCKIGDRLPKTFQHTNEVNAIKWPTGNLLASCDDMTLKIWSMKQDN 327
QY 383 CVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASAFDSTVRLMDVDRGICHTLTTHQ 442
DB 328 CVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASAFDSTVRLMDVDRGICHTLTTHQ 387
```

08-DEC-2000; 2000US-0251989P.  
08-DEC-2000; 2000US-0251990P.  
11-DEC-2000; 2000US-0254097P.  
05-JAN-2001; 2001US-0259678P.  
(HUMA-) HUMAN GENOME SCI INC.  
Rosen CA, Barash SC, Ruben SM;  
WPI; 2001-476159/51.  
N-PSDB; ADM19381.  
Isolated nucleic acid molecule encoding a channel/transporter protein is  
used in preventing, treating or ameliorating a medical condition.  
Claim 11; SEQ ID NO 667; 809pp; English.  
The invention relates to an isolated nucleic acid molecule encoding a  
channel/transporter protein or sequences at least 95% identical to a  
these. The nucleic acids and proteins encoded by them are used to  
prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
rabbits, goats, horses, cats, dogs, chickens or sheep. They are also used  
in diagnosing a pathological condition or susceptibility to a  
pathological condition. The antibodies to the proteins can also be used  
in alleviating symptoms associated with the disorders and in diagnostic  
immunoassays e.g. radioimmunoassays or enzyme linked immunosorbent assays  
(ELISA). Disorders which are diagnosed or treated include autoimmune  
diseases e.g. rheumatoid arthritis, hyperproliferative disorders e.g.  
neoplasms of the breast or liver, cardiovascular disorders e.g. cardiac  
arrest, cerebrovascular disorders e.g. cerebral ischemia, angiogenesis,  
nervous system disorders e.g. Alzheimer's disease, infections caused by  
bacteria, viruses and fungi and ocular disorders e.g. corneal infection.  
The polypeptides can also be used to aid wound healing and epithelial  
cell proliferation, to prevent skin aging due to sunburn, to maintain  
organs before transplantation, for supporting cell culture of primary  
tissues, to regenerate tissues and in chemotaxis. The polypeptides can  
also be used as a food additive or preservative to increase or decrease  
storage capabilities. This sequence corresponds to a protein of the  
invention.

Sequence 459 AA;  
Query Match 88.6%; Score 3250.7; DB 4; Length 459;  
Best Local Similarity 91.9%; Pred. No. 3.1e-48;  
Matches 452; Conservative 3; Mismatches 4; Indels 33; Gaps 3;

23 HSAFTFGESHISQSNINGALVPPAALISIIQKQYVEAEVSEINEDGTLFDRGPISLS 82  
1 H-A-----SD-----VEAEVSEINEDGTLFDRGPISLS 27

83 LIDAVMPDVVQTRQAYRDKLAQAAAAAATAATNQGSXKNGENTANGENGARTIA 142  
28 LIDAVMPDVVQTRQAYRDKLAQAAAAAATAATNQGSXKNGENTANGENGARTIA 87

143 NNHTDMNEVDGVEIPNKAVALRGHSEVFCANPVSDDLIVSGSDSTARIWNLSNS 202  
88 NNHTDMNEVDGVEIPNKAVALRGHSEVFCANPVSDDLIVSGSDSTARIWNLSNS 147

203 TSGPTQLVLRHCIRREGQDVPSNKNVTSLDWNSGTLATGSDGFGARITWKDGNLASTL 262  
148 TSGPTQLVLRHCIRREGQDVPSNKNVTSLDWNSGTLATGSDGFGARITWKDGNLASTL 207

263 GQHKGPFIALKWKNKGNFILSAGVDKTTIWDATGCAKQFPFHSAPALDWDVQSNNTF 322  
208 GQHKGPFIALKWKNKGNFILSAGVDKTTIWDATGCAKQFPFHSAPALDWDVQSNNTF 267

323 ASCSTDMCIHVCKIGDRLPKTFQHTNEVNAIKWPTGNLLASCDDMTLKIWSMKQDN 382  
268 ASCSTDMCIHVCKIGDRLPKTFQHTNEVNAIKWPTGNLLASCDDMTLKIWSMKQDN 327

383 CVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASAFDSTVRLMDVDRGICHTLTTHQ 442  
328 CVHDLQAHNKEIYTIKWSPTGPTNNPNANMLASAFDSTVRLMDVDRGICHTLTTHQ 387

443 EPVYSVAFSPDGRYLASGSDKCVHIWNTOTGALVHSYRGTTGGIFEVCMNAGDKVGCASA 502  
388 EPVYSVAFSPDGRYLASGSDKCVHIWNTOTGALVHSYRGTTGGIFEVCMNAGDKVGCASA 447  
503 SDGSVCVLDLRK 514  
448 SDGSVCVLDLRK 459

RESULT 9  
ABB60376  
ID ABB60376 standard; protein; 700 AA.  
XX  
AC ABB60376;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 7920.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US009231.  
PF  
XX 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PWD, Myers EW;  
PI  
XX WPI; 2001-656860/75.  
DR N-PSDB; ABL04479.  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 7920; 21pp + Sequence Listing; English.  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA  
CC sequences (ABLI01840-ABLI6175) and the encoded proteins (ABBS7737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 700 AA;  
Query Match 79.5%; Score 2916.9; DB 4; Length 700;  
Best Local Similarity 57.3%; Pred. No. 5.4e-41;  
Matches 411; Conservative 51; Mismatches 34; Indels 221; Gaps 61;

1 MSISDEVNLFVRYLQESGSHSAFTGIESHSIQSNINGALVPPAALISIIQKQYV 60  
1 MSFSDEVNLFVRYLQESGSHSAFTGIESHSIQSNINGALVPPAALISIIQKQYV 60  
61 EAEVSEINEDGTLFDRGPISLSLIDAMPDV-----VQTRQ-----AYPD---KLA-- 104  
61 EAEVSEINEDGTLFDRGPISLSLIDAMPDV-----VQTRQ-----AYPD---KLA-- 104  
61 EVESVGEDG---EVAPFIEGLSLIDAMPVEVKPLIVKT-EPGKPGAV-DSSAP-AGG 114  
105 -Q-----QH-----A-A-----A-----A-----AA---A-----A 115



Db 115 NQNNNAKPEIKBPGTGVAGSAGNKIAGSTTGTSTPTDQSASEVDSGNAANAGTVA 174  
QY 116 ---A---T---N-----Q-----QSAK--N-G-EN--T-- 130  
Db 175 GNNAGCQNAQSTGSGNSTPTPAGGDLAAPGASQKQNSNEAGSSSGNAGNANATSTD 234  
QY 131 -A-----NG-----EE-----N-----G-AH---TI- 141  
Db 235 AASSTSTNGSSSTSSVQEQPTSGLTGAGTGTSTNPDAASGASGATGSKAGAVTIR 294  
QY 142 --A--NN-----H-----T-DMVEY--DGD--VEIP-SNKAVLVR 166  
Db 295 VQAQGNVQSGSSNAOSSAPSGTISSTSGGAGTPAAL-VPMIDENIEIPES-KARVLR 352  
QY 167 GHESEVFCIAWNPV--DLVSGSGDSTARIWNLE-NSTSGTQVLVLRHCIRGGQDVPS 224  
Db 353 GHESEVFCIAWNP-SRDLASGSGDSTARIWDSANTNS--NQLVLRHCIRGGQDVPS 409  
QY 225 NKDVTSLDWNSEGTLLATGSDGFARIW-TKQGNLSTLQHGKPIFALKWKK-GNFIL 282  
Db 410 NKDVTSLDWNCDGSLATGSDGYARIWKT-DGRLASTLQHGKPIFALKW-KCNGYIL 467  
QY 283 SAGVDKTTIWDHATGE-AKQPPFHSAPALVDVDSNNNTFASCSTD-MCIHVCKLG--Q 338  
Db 468 SAGVDKTTIWDASTGQCT-QEAFHSAPALVDVQTNQAFASCSTDQ--IHVCRILGVNE 525  
QY 339 DRPIKTFQHTNEVNAIKWDPKTNLLASCSDMTLKIWSMKQD-NCVHDLQAHNKEIYTI 397  
Db 526 --PIKTFKHTNEVNAIKWCPQQLASCSDMTLKIWSMNRDCC-HDLQAHNKEIYTI 582  
QY 398 KWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTKHQBPVYSVAFSPDGRYL 457  
Db 583 KWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTKHQBPVYSVAFSPDGRYL 642  
QY 458 ASGSFDPKCVHIWNTQCALVHSYRGTCGIPEVCWNAAGDKVGSASDGSVCVLDLRK 514  
Db 643 ASGSFDPKCVHIWNTQCALVHSYRGTCGIPEVCWNSKGTGVGSASDGSVCVLDLRK 599

RESULT 10  
ABG21351  
ID ABG21351 standard; protein; 584 AA.  
XX AC ABG21351;  
XX DT 18-FEB-2002 (first entry)  
XX DE Novel human diagnostic protein #21342.  
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX KW food supplement; medical imaging; diagnostic; genetic disorder.  
XX OS Homo sapiens.  
XX PN WO200175067-A2.  
XX XX WO200175067-A2.  
XX PD 11-OCT-2001.  
XX PF 30-MAR-2001; 2001WO-US008631.  
XX PR 31-MAR-2000; 2000US-00540217.  
XX PR 23-AUG-2000; 2000US-00649167.  
XX XX (HYSE-) HYSEQ INC.  
XX XX Drmanac RT, Liu C, Tang YT;  
XX PI WPI; 2001-639362/73.  
XX DR N-PSDB; AAS85538.  
XX DR New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess  
XX biodiversity.  
XX Claim 20; SEQ ID NO 51710; 103pp; English.  
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)  
CC sequences. (I) is useful as hybridisation probes, polymerase chain  
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,  
CC and in recombinant production of (II). The polynucleotides are also used  
CC in diagnostics as expressed sequence tags for identifying expressed  
CC genes. (I) is useful in gene therapy techniques to restore normal  
CC activity of (II) or to treat disease states involving (II). (II) is  
CC useful for generating antibodies against it, detecting or quantitating a  
CC polypeptide in tissue, as molecular weight markers and as a food  
CC supplement. (II) and its binding partners are useful in medical imaging  
CC of sites expressing (II). (I) and (II) are useful for treating disorders  
CC involving aberrant protein expression or biological activities. The  
CC polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic  
CC amino acid sequences. The patent did not appear in the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX SQ Sequence 584 AA;  
Query Match 76.1%; Score 2791.1; DB 4; Length 584;  
Best Local Similarity 63.1%; Pred. No. 5.6e-39;  
Matches 383; Conservative 44; Mismatches 41; Indels 139; Gaps 52;  
QY 2 SI-----S--S-----D-----EV---N---FLV-YRY-LQE-SCFSSHSAFTGIESH 33  
Db 22 AIGAVGFRQAASLAGQPDGNGVCVHAGNSRIF-VPLPVGL-EVKGFSHSAFTGIESH 79  
QY 34 ISOSNINGALVPPAALISIIQKGLQYVEAEVSNEDGTLFDGRIPIESLSLIDAVMPDVQ 93  
Db 80 ISOSNINGTLVPPSALISILQKGLQYVEAEISINKDGTVDSPRIESLSLIVAVPDVQ 139  
QY 94 TQQAYRDKLAQQA--AAA--A--AAT-----NQGSAAKNGENTANGAHTIA 142  
Db 140 MRQQAPEKLTQQAASAAATEASAMAKAATMTPAALISQNPVKREATVNGEAGAHEI- 198  
QY 143 NNHTDM-MEVDGDVEIPSNKAVVLRGHESEVFCIAWNPVSDLLVSGSGDSTARIWNLSEN 201  
Db 199 NNHS-KPMEIDGDVEIPPNKATVLRGHESEVFCIAWNPVSDLLVSGSGDSTARIWNLSEN 257  
QY 202 STSGPTQLVLRHCIRREGQDVPSNKDVTSLDWNSEGTLLATGSDGFARIWTKDGNLST 261  
Db 258 SNGGSTQLVLRHCIRREGHDVPSNKDVTSLDWNSDGTLAMGSDYDGFARIWTE--N-A-- 312  
QY 262 LQGHKGPFIAP--KW-NKXGN--FILS-AGVDKTTIWDHAT--GEAKQPPFHSAPAL 312  
Db 313 -----P--ALDWDWQ--NMTFA-SCS-TD--CI---HVCRLGCD-----H--P-- 344  
QY 313 DVD-WQSNNTFASCSTDMDCIH-VC-----KL-----G-QDRPIKTF-----QG- 347  
Db 345 -VKTFQ-GHTF--C-T--CIESICFWGGLRKLTTMTTEGKRLRP-KTFCSDGGALLPPAGR 396  
QY 348 --H--T-----NEVNAIKWDPKTNLLASCSDMTLKIWSMKQDNCVHDLQA 389  
Db 397 RPHLLTGPDIFPKRLSALFQNEVNAIKWDPKTNLLASCSDMTLKIWSMKQDNCVHDLQA 456  
QY 390 HNKEIYTIKWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTKHQBPVYSVA 449  
Db 457 HSKEIYTIKWSPTGPGTNNPNANMLASASFDSTVRLWDVDRGICHTLTKHQBPVYSVA 516  
QY 450 FSPDGRYLASGSDKCVHIWNTQCALVHSYRGTCGIPEVCWNAAGDKVGSASDGSVCV 509  
Db 517 FSPDGRYLASGSDKCVHIWNTQCALVHSYRGTCGIPEVCWNAAGDKVGSASDGSVCV 576

QY 510 L-DLR-K 514  
 Db 577 LWLQSGK 583

RESULT 11  
 ID ABP51424  
 XX ABP51424 standard; protein; 395 AA.  
 AC ABP51424;  
 XX  
 DT 03-SEP-2002 (first entry)  
 XX  
 DE Human MDDT SEQ ID NO 446.  
 XX  
 KW Human; MDDT; disease detection and treatment molecule polynucleotide;  
 KW proliferative disorder; hepatitis; psoriasis; cancer; AIDS;  
 KW autoimmune disorder; inflammatory disorder; allergy; multiple sclerosis;  
 KW rheumatoid arthritis; transgenic; gene therapy; antiarteriosclerotic;  
 KW hepatotropic; antiinflammatory; antipsoriatic; cytosclerotic; anti-HIV;  
 KW antiallergic; antianaemic; antiaethmatic; antiatherosclerotic; antitigout;  
 KW neuroprotective; antirheumatic; antiarthritic.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200240715-A2.  
 XX  
 PD 23-MAY-2002.  
 XX  
 PF 06-SEP-2001; 2001WO-US027628.  
 XX  
 PR 05-SEP-2000; 2000US-0229747P.  
 PR 05-SEP-2000; 2000US-0229748P.  
 PR 05-SEP-2000; 2000US-0229749P.  
 PR 05-SEP-2000; 2000US-0229750P.  
 PR 05-SEP-2000; 2000US-0229751P.  
 PR 05-SEP-2000; 2000US-0230583P.  
 PR 06-SEP-2000; 2000US-0230505P.  
 PR 06-SEP-2000; 2000US-0230514P.  
 PR 06-SEP-2000; 2000US-0230515P.  
 PR 06-SEP-2000; 2000US-0230517P.  
 PR 06-SEP-2000; 2000US-0230518P.  
 PR 06-SEP-2000; 2000US-0230519P.  
 PR 06-SEP-2000; 2000US-0230595P.  
 PR 06-SEP-2000; 2000US-0230597P.  
 PR 06-SEP-2000; 2000US-0230598P.  
 PR 06-SEP-2000; 2000US-0230599P.  
 PR 06-SEP-2000; 2000US-0230610P.  
 PR 06-SEP-2000; 2000US-0230611P.  
 PR 06-SEP-2000; 2000US-0230865P.  
 PR 06-SEP-2000; 2000US-0230988P.  
 PR 06-SEP-2000; 2000US-0230989P.  
 PR 07-SEP-2000; 2000US-0230951P.  
 PR 07-SEP-2000; 2000US-0231163P.  
 PR 07-SEP-2000; 2000US-0231167P.  
 XX  
 PA (INCY-) INCYTE GENOMICS INC.  
 XX  
 XX Jackson S, Lincoln SE, Altus CM, Dufour GE, Chalup MS;  
 PI Hillman JL, Jones AL, Yu JT, Wright RJ, Gietzen D, Liu TP, Yap PE;  
 PI Dahl CR, Momiyama MG, Bradley DL, Rohatgi SD, Harris B;  
 PI Roseberry AM, Gerstin EH, Peralta CH, David MH, Panzer SR, Flores V;  
 PI Daffo A, Marwaha R, Chen AD, Chang SC, Au AP, Inman RR;  
 XX WPI: 2002-527544/56.  
 DR N-PSDB; ABQ72641.  
 DR  
 XX Novel human disease detection and treatment polypeptide, useful in  
 PT diagnosis, prevention or treatment of cell proliferative disorders e.g.  
 PT arteriosclerosis, cirrhosis and an autoimmune/inflammatory disorder e.g.  
 PT AIDS.  
 XX  
 PS Claim 14; Page 578; 618pp; English.  
 XX

CC The invention relates to an isolated human disease detection and  
 CC treatment (MDDT) polypeptide (I) selected from a polypeptide having a  
 CC sequence selected from 254 sequences (ABP51231-ABP51484) given in the  
 CC specification, a naturally occurring polypeptide comprising a sequence  
 CC having at least 90% identity to (I) or a biologically active or  
 CC immunogenic fragment of (I). (I) is useful for screening a compound for  
 CC effectiveness as an agonist or antagonist, for screening a compound that  
 CC specifically binds (I) or modulates the activity of (I), and for  
 CC preparing a polyclonal or monoclonal antibody by hybridoma technology.  
 CC Nucleic acids (II) (ABQ72449-ABQ72700) encoding (I) are useful for  
 CC screening a compound for effectiveness in altering expression of a target  
 CC polynucleotide comprising. Oligonucleotides and antibodies are useful for  
 CC detecting MDDT in a sample or for assessing toxicity of a test compound,  
 CC in a diagnostic test for a condition or a disease associated with the  
 CC expression of MDDT in a biological sample, for detecting (I) in a sample,  
 CC and for purifying (I) from a sample. A composition comprising (I), an  
 CC agonist or antagonist is useful for treating a disease or condition  
 CC associated with decreased or increased expression of functional MDDT. (I)  
 CC or (II) are useful for diagnosing, treating or preventing disorders  
 CC associated with aberrant expression of MDDT, where the disorders are  
 CC selected from a cell proliferative disorder such as arteriosclerosis,  
 CC cirrhosis, hepatitis, psoriasis, and cancer and an  
 CC autoimmune/inflammatory disorder such as AIDS, Addison's disease,  
 CC allergy, anaemia, asthma, atherosclerosis, gout, multiple sclerosis or  
 CC rheumatoid arthritis. (II) are useful for creating knockin humanised  
 CC animals or transgenic animals to model human diseases, in somatic or  
 CC germline gene therapy, to generate a transcript image of a tissue or cell  
 CC type, for detecting differences in the chromosomal location due to  
 CC translocation or inversion among normal, carrier or affected individuals  
 CC and as hybridisation probes for mapping naturally occurring genomic  
 CC sequences  
 XX  
 SQ Sequence 395 AA;

Query Match 74.0%; Score 2716; DB 5; Length 395;  
 Best Local Similarity 98.5%; Pred. No. 2.3e-38;  
 Matches 385; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSISDEVNPLVYRYLQSGFSGHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 60  
 Db |||||  
 QY 5 MSISDEVNPLVYRYLQSGFSGHSAFTGIESHISQSNINGALVPPAALISIIQKGLQYV 64  
 Db |||||  
 QY 61 EAEVSNEDGTLFDGRPIESLSIDAAMPDVVQTRQOAVRDKLAQHAAAAAATAATNQ 120  
 Db |||||  
 QY 65 EAEVSNEDGTLFDGRPIESLSIDAAMPDVVQTRQOAVRDKLAQHAAAAAATAATNQ 124  
 Db |||||  
 QY 121 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGDVEIPSNKAVLVRGHESEVFICAWNPFV 180  
 Db |||||  
 QY 125 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGDVEIPSNKAVLVRGHESEVFICAWNPFV 184  
 Db |||||  
 QY 181 SDLLVSGSGDSTARINWLNSENSTGPTQLVRHCIREGGQDVPNSKDVTSLDWNSGTL 240  
 Db |||||  
 QY 185 SDLLVSGSGDSTARINWLNSENSTGPTQLVRHCIREGGQDVPNSKDVTSLDWNSGTL 244  
 Db |||||  
 QY 241 ATGSYDGPARIWTKDGNLASTLGQHKGPFAFKWKNKGNFILSAGVDKTTIWDHAHTGEA 300  
 Db |||||  
 QY 245 ATGSYDGPARIWTKDGNLASTLGQHKGPFAFKWKNKGNFILSAGVDKTTIWDHAHTGEA 304  
 Db |||||  
 QY 301 KQOFPFHSAPALDVQSNNTFASCTDMCIHVCKLQGDPRPKTFQGHNEVNAIKWDPT 360  
 Db |||||  
 QY 305 KQOFPFHSAPALDVQSNNTFASCTDMCIHVCKLQGDPRPKTFQGHNEVNAIKWDPT 364  
 Db |||||  
 QY 361 GNLASCSDDMTLKIWSMKQDNCVHDLQAHN 391  
 Db |||||  
 QY 365 GNLASCSDDMTLKIWSMKQDNCVHDLQAHN 395

RESULT 12  
 ID ABP60991  
 XX ABP60991 standard; protein; 5635 AA.  
 AC ABP60991;  
 XX

DT 10-SEP-2002 (first entry)  
 XX Novel human protein. SEQ ID 78.  
 XX  
 KW Human; cytostatic; vulnary; antiarteriosclerotic; antiparkinsonian;  
 KW neurotropic; neuroprotective; immunosuppressive; haemostatic;  
 KW antiinflammatory; cardiant; antitumor; virucide; antithyroid;  
 KW cerebroprotective; anorectic; metabolic; vaccine; cancer; infection;  
 KW wound healing disorders; atherosclerosis; Parkinson's disease;  
 KW Alzheimer's disease; autoimmune disorder; haematopoietic disorder;  
 KW inflammation; neoplastic disease; nervous system disorder;  
 KW cardiovascular disorders; pancreatitis; respiratory disorder;  
 KW hyperproliferation; systemic autoimmune disease; hyper-immunity;  
 KW developmental abnormality; gastrointestinal ulceration; neuropathy;  
 KW haematological disease; metabolic disease; sperm dysfunction;  
 KW thyroid disorder; hypothyroidism; brain damage; colitis;  
 KW cone photo- transduction deficiency; neurological disease; stroke;  
 KW angiogenesis; ovulation disorder; spinal cord; thyroid gland; heart;  
 KW trachea; thymus; lymph node; muscular system; obesity; anorexia;  
 KW growth abnormality; precocious puberty.  
 XX Homo sapiens.  
 XX WO200250105-A1.  
 XX  
 XX 27-JUN-2002.  
 XX  
 XX 17-DEC-2001; 2001WO-US049232.  
 XX  
 XX 19-DEC-2000; 2000US-0256710P.  
 XX 20-DEC-2000; 2000US-0257048P.  
 XX 09-JAN-2001; 2001US-0260482P.  
 XX 30-JAN-2001; 2001US-0264922P.  
 XX 06-FEB-2001; 2001US-0266797P.  
 XX 19-MAR-2001; 2001US-0276988P.  
 XX 04-APR-2001; 2001US-0281535P.  
 XX 08-MAY-2001; 2001US-0289622P.  
 XX (SMIK ) SMITHKLINE BEECHAM CORP.  
 XX (SMIK ) SMITHKLINE BEECHAM PLC.  
 XX (GLAX ) GLAXO GROUP LTD.  
 XX  
 XX Agtarwal P, Birkeland M, Cogswell JP, Kabnick KF, Lai Y;  
 XX Martensen SA, Rizvi SK, Smith RF, Strum JC, Xie Q;  
 XX WPI: 2002-508784/54.  
 XX N-PSDB; ABQ86156.  
 XX  
 XX Secreted proteins and polynucleotides useful as vaccines for preventing  
 XX or treating various diseases e.g. cancer, wounds, atherosclerosis,  
 XX Parkinson's disease, Alzheimer's disease, infection, autoimmune disorder.  
 XX Claim 1(a); Page 285-297; 335pp; English.  
 XX  
 XX The invention relates to an isolated polypeptide with signal sequences  
 XX which allow it to be secreted extracellularly or membrane associated. The  
 XX activity of polypeptides of the invention may be described as,  
 XX cytostatic, vulnary, antiarteriosclerotic, antiparkinsonian, neurotropic,  
 XX neuroprotective, immunosuppressive, haemostatic, antiinflammatory,  
 XX cardiant, antitumor, virucide, antithyroid, cerebroprotective, anorectic,  
 XX and metabolic. Polypeptides and polynucleotides of the invention are  
 XX useful in the treatment, or as a vaccine in the prevention of, cancer,  
 XX wound healing disorders, infection, atherosclerosis, Parkinson's disease,  
 XX and Alzheimer's disease, autoimmune disorder, haematopoietic disorder,  
 XX inflammation, neoplastic diseases, nervous system related disorders and  
 XX cardiovascular disorders, pancreatitis, respiratory disorder,  
 XX hyperproliferation, systemic autoimmune disease, hyper-immunity,  
 XX developmental abnormality, gastrointestinal ulceration, neuropathy,  
 XX haematological diseases, metabolic diseases, sperm dysfunction, thyroid  
 XX disorders e.g. hypothyroidism, brain damages, colitis, cone photo-  
 XX transduction deficiency, neurological diseases, stroke, angiogenesis,  
 XX ovulation disorders, diseases in the spinal cord, thyroid gland, heart,  
 XX trachea, thymus, lymph node and muscular system, obesity, anorexia,

CC growth abnormalities, and alleviation of precocious puberty. The  
 CC sequences given in records ABP60965-ABP61019 represent novel human  
 CC proteins of the invention  
 XX  
 XX Sequence 5635 AA;  
 Query Match 71.7%; Score 2631.9; DB 5; Length 5635;  
 Best Local Similarity 9.6%; Pred. No. 3.3e-31;  
 Matches 440; Conservative 57; Mismatches 15; Indels 4071; Gaps 372;  
 QY 1 M-S-----I-----SS--D-----EV----- 8  
 DB 1 MISWEVHTVFLFALLYSLAQDASQPSIRAEIEPEGASTLAFVFDVTGSMYDDLQVVI 60  
 QY 9 -----NF-LV----- 12  
 DB 61 EGASKILETSLKRPRKPLNFALVPPHDPPEIGPVITTPDKPKFQYELRELYVGGDCPE 120  
 QY 13 -----Y-----R-----Y-----LQ----- 17  
 DB 121 MSIGAIIKIALEISLPGSFIYVFTDARS KDYRLTHEVLQLIQKQSQVFLVTGDCDRTH 180  
 QY 18 -----E-----SG--P-----S--H--S-----A-----F 26  
 DB 181 IGYKVVEIASTSSGQVFLDKKQVNEVLKQVVEAAVQASKVHLLSTDHLEQAVNTWRIPF 240  
 QY 27 -----T-----FG-----I-----E----- 31  
 DB 241 DPSLKEVTLSLGPSPMIEIRNPLKLIKKGLGHELLHINSKVVNVKEPEAGMWTVK 300  
 QY 32 -----S--H-----I-----S--Q----- 36  
 DB 301 TSSSGRHSVRITGLSTIDFRAGSRKPTLDFKKTYSRPQGIPTVYLLNTSGISTPARID 360  
 QY 37 -----S-----NI-----N-----G----- 41  
 DB 361 LLELLISGSLKTIPTVKYTPHRYKPYGIWNISDFVPPNEAFFLKVTGYDKDDYLFQVSS 420  
 QY 42 -----A-----L-----VP-----P-----A 47  
 DB 421 VSFSSIVDPAPKVTMPKTPGGYLLQPGQIPCSVDLSLLPFTLSFVRNGVTGLVDQYLKESA 480  
 QY 48 -----A-----L-----I-----S-----IIQ-----KG----- 56  
 DB 481 SVNLDIAKVTLSDEGFYECIAVSSAGTGRAQTFDFVSEPPPIQVNNVTVTPGERAVLT 540  
 QY 57 -----L--Q-----Y-----V--B----- 61  
 DB 541 CLIIISAVDYNLTQWRNDRDLRLAEPARTLANLSLELKVKNFNDAGEYHCVSSEGGSS 600  
 QY 62 -A-----EVS-I-----NE----- 68  
 DB 601 AASVFLTVQEPKVTMPKQKQSFQSGSEVSMCSATGYKPKKIATVNDMFIVGSHRYM 660  
 QY 69 --DGLTF-----D-----G-----R-----P-----I-- 78  
 DB 661 TSDGTLFIKNAAPKAGIYVGLASAGTDKQNSTLYIEAPKLMVQSELLVALGDITV 720  
 QY 79 -B--S-----L--S--LI-----DA-----V----- 87  
 DB 721 MECKTSGIPPPQVQKFGDLRLPSTFLIDPLLLGLLKIQTQDLADAGDYTCVAINAAGR 780  
 QY 88 -----M-----P-DV-----VQ-----T----- 94  
 DB 781 ATGKITLDVGSPPVFIQEPADVMEIGSNVTLPVYVQGYPEPTIKWRRLDNMPISRPFS 840  
 QY 95 -----R----- 95  
 DB 841 VSSIISOLRTGALFILNLWASDKGTICEAENQFGIKQSETTVTGTGLVAPLIGISVAN 900  
 QY 96 -----QQ-----A-----Y-----R-D-KL-----AQ----- 105  
 DB 901 VIEGQQLTLPCLLAGNPITPERRWIKNSAMLLQNPNYITVRS DGLS LHIERVQLQDGGEYTC 960

QY 106 -----OH-----A----- 108  
Db 961 VASVAGTNNKTTVVVHVLPTIOHQOILSTIEGIPVTLPCASGNPKPSVWSKKGEL 1020  
QY 109 -----A--A--A-----A--A--A----- 114  
Db 1021 ISTSAKESAGADGSLYVSPGEBSEYVCTATNTAGYAKVKQLTVYVPRVFGDQRG 1080  
QY 115 -----A-----A----- 116  
Db 1081 LSQDPVBISVLAGEBVLPCVKSLPPPIITWAKETOLISPPSPRHTFLPSGSMKITET 1140  
QY 117 -----ATN-----Q-----Q--GA--K-----N----- 126  
Db 1141 RTSDSGMYLCVATNIAGNVTQAVKLVNHPVKIORGPXKLVQVQORVDIPCAQGTPLP 1200  
QY 127 -----GE--N-----T--A--N--G----- 133  
Db 1201 VITWSKGSWLVDEGHHVSNPDGTLSDIQAATPSDAGIYTCVATNIAGTDETEITLHVQE 1260  
QY 134 -----E--N-----GA-----H----- 139  
Db 1261 PPTVEDLEPPYNTTFOERVANORIEFPCPAKGTPTIKWHLNGRELTCRGPGISILEDG 1320  
QY 140 T--IA-----N-----N--HT--DM----- 148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHPVPIKBEQVTVNSVLLNQLTNL 1380  
QY 149 -----M-----E----- 150  
Db 1381 PCEVEGTPSPIMMYKDNVQVTESTTIQTVNNGKILKILFRATPEDAGRYSCAKINIAGTS 1440  
QY 151 -----VD-----G-----DV--B-----I-----P-- 158  
Db 1441 QKYPNIDVLPPTTIIGNFNEVSVLNRDVALEQVKGTPFPDIHMFKDGLFLGDPN 1500  
QY 159 -----SN--K--A-----V-----VL 165  
Db 1501 VELDROVLHLKNARENDKRYOCTVSNAGKQAKDKILTIYNPSPKGNVTTDISVL 1560  
QY 166 -----RG-----H----- 168  
Db 1561 INSLIKETETRLPMPAITWKDQBPIMSSQALYIDKQYLHI PRAQVSDSATYTVCHV 1620  
QY 169 -----E-S-----E-----VFI-----C--A-----W----- 177  
Db 1621 ANVAGTAEKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECKAAGNPSPILTWLDGV 1680  
QY 178 -----N-----P--V--SD-- 182  
Db 1681 PVKANDNIRIEAGGKLEINSAQEBIDRGQYICVATSVAGEKEIKEYVDVLVPPAIEGGDE 1740  
QY 183 -----LL--VSGS-----G-----D-----S-- 191  
Db 1741 TSYFVIMVNNLLELDCHVTGSPPTIMWLKDLIDERDGFILLNGRKLVIQAQVSN 1800  
QY 192 -----TA-----R-----I--W 196  
Db 1801 GLYRCMAANTAGDHKEFEVTVHVPPTIKSSGLSERVVVKYKPVALQCTIANGIPNPSITW 1860  
QY 197 -----NL--S-----E-----NS--T-----S 204  
Db 1861 LKDDQPNTAQGNLKIQQSGRVLQIAKLTLEDAGRYTCVATNAAGETQQHIQLHVHEPPS 1920  
QY 205 -----G-----PTOL-----V-----L--R----- 212  
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVEVITWYKDNLLSGSTSMTFLNRQIIDI 1980  
QY 213 -----H-----C-----I-- 215  
Db 1981 AQISDAGYKCVAINAGATFELFYSIQVHVAPSIGSNMNVAVVNNPVRLSEARGIPA 2040

QY 216 -----R-----E--G-----G--Q--D-- 221  
Db 2041 PSLTWLKDGPVSSFSNGLQVLGGRIALTSAQISDGTGRYTCVAVNAAGEKQDIDLVR 2100  
QY 222 -VPSN-----KD-----VT--S--L-- 231  
Db 2101 YVFPNMBEQNVSVLISQAVELLCSDAIPPTLTWLDKGHLLKPKGLSISENSVLK 2160  
QY 232 -----D-----W--N--S-----EG--TLL--A 241  
Db 2161 IEDAQVDTGRYTCVATNAGTEKNYVNVNWPPIGSGDELQTLTVIEGNLISLCE 2220  
QY 242 T-----GS--Y--D-----G-----F-----A-- 249  
Db 2221 SGIPPPNLIWKKGSPVLTDMSGRVRLSGGROLQISIAEKSDAALYSCVAVNAGTAK 2280  
QY 250 -----R-----I-----I-----WKDG-- 256  
Db 2281 EYNLQVIRPTITNSGSHPTIIVTRGKSISLECEVOGIPPTVTWMDKHPLIKAGVE 2340  
QY 257 -----NL--AS-----T--L-----GOH-- 265  
Db 2341 ILDEGHILQIKNIHVSDTGRYVCVAVNAGMTDKKYDLSVHAPPSIIGNHRSNENISVVE 2400  
QY 266 K-----G--P--I--FALK--W----- 274  
Db 2401 KNSVSLTCEASGIPLSITWF--KDGWPVSLNSVRILSGRMLRLMOTTMEDAGQYTCV 2458  
QY 275 -----N-----K--K-----GN----- 279  
Db 2459 VRNAGEERKI FGLSVLVPPIHIVGENTLEDVKVKEKQSVLTICEVTGNPVPITWHKQOQ 2518  
QY 280 -----F--I-----L-----S--AGV-- 286  
Db 2519 PLQEDEAHHIISGGRFLQITNVQVPHGTGRYTCCLASSPAGHKRSFSLNVFVPTIAGVGS 2578  
QY 287 D--K--T-----T--I-----I--W----- 293  
Db 2579 DGNPVDVILNSPTSLVCEAYSPPATITWPKOGTPLESNRNIRILPGGRTLQILNAQE 2638  
QY 294 D-----A--H-----TG-----BA 300  
Db 2639 DNAGRYCVATNEAGEMIKHYEVKYVYIPIINKDGLWGPGLSPKEVKIKVNNLTILECA 2698  
QY 301 -----K-----Q--Q-----F 304  
Db 2699 YAIPASLSWYKDGQPLKSDDHVNIANGHTLQIKEAQISDTGRYTCVASNIAGEDELDF 2758  
QY 305 -----P--F-----H-- 307  
Db 2759 DVNIQVPPSFQKLWEIGNWMLDTCRNGEAKDVIIINNPIISLYCETNAAPPPTLTWYKDGHPL 2818  
QY 308 -SA-----P-----A-----LDV-----D-- 315  
Db 2819 TSSDKVLILPGGRVLIQIPRAKVEDAGRYTCVAVNEAGDSQYDVRVLVPPPIKANGS 2878  
QY 316 -----WQ-----SN----- 319  
Db 2879 PEEVTVLNVKSALIECLSSGSPAPRNSWQDQGPILLEDHKKFLSNGRILQILNTQITDI 2938  
QY 320 -----NT-----FAS--C-----S-- 326  
Db 2939 GRVYVCAENTAGSACKYFNLVNHPVPSVIGPKSENLTVVVNNFISLTCEVSGFPDLSW 2998  
QY 327 -----T-----D--MCI----- 331  
Db 2999 LKNEQPIKINTNLIVPGGRTLQIIRAKVSDGGETCIAINQAGESKKKFSILTVVPPSI 3058  
QY 332 --H-----V--C-----K-----L--QO----- 338  
Db 3059 KDHSSESLSVNVNREGTSVLECESNAVPPPVITWYKGRMITESTHVIELADGOMLHIK 3118  
QY 339 -----D-----R-----P--IK-----TF----- 345



CC for therapeutic intervention in treating a disease associated with  
 CC altered mitochondrial function. Specifically, it refers to a method for  
 CC identifying proteins of the human heart mitochondrial proteome that are  
 CC useful for drug screening assays, as well as therapeutic targets. The  
 CC present invention describes a method for identifying such proteins that  
 CC can be used in the treatment of various diseases associated with altered  
 CC mitochondrial function including diabetes mellitus, Huntington's disease,  
 CC osteoarthritis, Leber's hereditary optic neuropathy (LHON) mitochondrial  
 CC encephalopathy lactic acidosis and stroke (MELAS), myoclonic epilepsy  
 CC ragged red fibre syndrome (MERRF) or cancer. Accordingly, these  
 CC compositions have neuroprotective, neurotropic, antidiabetic,  
 CC anticonvulsant, antiarthritic, osteopathic, ophthalmological and  
 CC cyostatic activities. This polypeptide sequence is a human heart  
 CC mitochondrial protein of the invention.

XX  
 SQ Sequence 5636 AA;

Query Match 71.6%; Score 2627.8; DB 7; Length 5636;

Best Local Similarity 9.6%; Pred. No. 3.9e-31;

Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;

QY 1 M-S-----I-----SS--D-----EV-----8  
 DB 1 MISWEVHTVFLFALLYSSLAQDASQSEIRAEFFPEGASTLAFVFDVTGMYDDLQVVI 60  
 QY 9 -----NF-LV-----12  
 DB 61 EGASKILETSLKRPKRLFNALVPHDPEIGPVITTDPKFKQYELRELYVOGGDCPE 120  
 QY 13 -----Y-----R-----Y-----LQ-----17  
 DB 121 MSIGAIALEISLPGSFIVFTDARSKDYRLTHEVLQLIQKQSQVVFVLTGDCDDRTH 180  
 QY 18 -----E-----SG--F-----S--H--S-----A-----F 26  
 DB 181 IGYKVEEIASTSGQVHLDKQVNEVLKWEAVQAKVHLLSTDLHLQAVNTWRIFP 240  
 QY 27 -----T-----T-----FG--I-----E-----31  
 DB 241 DPSLKEVTVSLSGFSPMIEIRNPLGKLIKGFGLHLLNTHNSAKVVNVKEPEAGMWTVK 300  
 QY 32 ---S--H-----I-----S--Q-----36  
 DB 301 TSSSGRHSVRITGLSTIDFRAGFSRPTLDFKKTVRPVGQIPTVYLLNTSGISTPARID 360  
 QY 37 -----S-----S-----NI-----N-----G-----41  
 DB 361 LLELLSIGSSLKTI PVKYYPHRKPVGIMNISDFVPPNEAFFLKVTDYKDYLFORVSS 420  
 QY 42 -----A-----L-----VP-----P-----A 47  
 DB 421 VSFSSIVDPAPKVTMEKTPGYVLQPCQIPCSVDLSLLPFTLSFVRNGVTGLVDQYLKESA 480  
 QY 48 -----A-----L-----I-----S-----IIO-----KG-----56  
 DB 481 SVSLDIAKVTLSDEGFVECIASVAGSAGTQAFDFVSEPPPVQVFNNTVTGERAVLT 540  
 QY 57 -----L--Q-----L-----Y-----V--E-----61  
 DB 541 CLIIASVDYNLTWQRDRDVRLEAPARIPTLANLSLELKSVPKNDAGEYHCWVSSEGGSS 600  
 QY 62 -A-----EVSII-----NE-----68  
 DB 601 AASVELTVQBPVKVMPKQSTGGSEVSMCSATCYPKPKIAWTNDMFIVGSHRYRM 660  
 QY 69 --DGLTF-----D-----G-----R-----P-----I--78  
 DB 661 TSDGTLFIKNAKPDAGIYGLASNSAGTDKQNSTLRYIEAPKLMVQSELLVALGDITV 720  
 QY 79 -E--S-----L--S--LI-----L-----DA-----V-----87  
 DB 721 MECKTSGIPPPQVKNFKGDLRLPSTFLIIDPLLLGLLKIQETQDLQDGYTCVAINAEGR 780

QY 88 -----M-----P-DV-----VQ-----T-----94  
 DB 781 ATGKIITLDVGSPPVFIQEPADWSMEIGSNVTLCYVQGYPEPTIKWRLDNMFISRPFS 840  
 QY 95 -----R-----95  
 DB 841 VSSISQLRTGALFILNWSADKGTIYCEARNQFKIQSETTVTVTGLVAPLIGISPSVAN 900  
 QY 96 ---Q-----Q-----A-----Y-----R-D-KL-----AQ-----105  
 DB 901 VIEGQQLTLPCTLLAGNPPIPERWRKNSAMLLQNPYITVRSDGSLHIERVOLQDGGSEYTC 960  
 QY 106 -----QH-----A-----108  
 DB 961 VASNVAGTNNKTTISVVVHVLPITIQHQQLLSTIEGIPVTLPCKASGNPKPSVIMSKGEL 1020  
 QY 109 -----A--A--A-----A--A--A-----114  
 DB 1021 ISTSSAKFSAGADGSLVWVSPEGESGEYVCTATNTAGVAKRKVQLTVVYVRPVFGDLRG 1080  
 QY 115 -----A-----A-----116  
 DB 1081 LSQDKPVEISVLAGBEVTLPCFVKSLPPPIITWAKETQLISPPSPRHTFLPSSGMKITET 1140  
 QY 117 -----ATN-----Q-----Q-GSA--K-----N-----126  
 DB 1141 RTSDSGMVLCVATNIAGNVTQAVKLVNVHVPVKIQRPKHLKVQVQORVDIPCAQAQTPLP 1200  
 QY 127 -----GE--N-----T--A--N--G-----133  
 DB 1201 VITWSKGGSTMLVDGSHVSNPDGTLSDQATPSDAGIVTCVATNIAGTDETEITLHVQE 1260  
 QY 134 -----E-----E--N-----GA-----H-----139  
 DB 1261 PPTVELEPPYNTTFQERVANQRIEPPCPAKPTKTIKWLHNGRELGTREPGISILEEG 1320  
 QY 140 T--IA-----N-----N--HT-----DM-----148  
 DB 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKEQVSNVSVLLNQLTNL 1380  
 QY 149 -----M-----E-----150  
 DB 1381 FCEVEGTSPPIIMWYKDNQVVTESITQTVNNGKILKLFATPEDAGRYSCKAINIAGTS 1440  
 QY 151 -----VD-----G-----DV--E-----I-----P--158  
 DB 1441 QKYFNIDVLVPPTIIGTNFPKEVSVVLRNDRDALECOVKGTPTPPDIHWFKDGKPLFLGDPN 1500  
 QY 159 -----SN--K-A-----V-----VL 165  
 DB 1501 VELLDRGQVLHLKNARRNDKGRVQCTVSNAAAGKQAKDIKTIYIPPSIKGNVTDISVL 1560  
 QY 166 -----RG-----H--168  
 DB 1561 INSILKLECKTRGLPMAITWYKDGQIPMSSSQALYIDRGQYLHI PRAQVSDSATYTCV 1620  
 QY 169 -----E-S-----E-----VFI-----C--A-----W-----177  
 DB 1621 ANVAGTAESFHVVDVVPVPMIEGNLATPLNKQVVIASHTLECNAAAGNSPILTWLKDGV 1680  
 QY 178 -----N-----P-V--SD--182  
 DB 1681 PVKANDNFRIEAGGKLEITMSAQEIDRGQYICVATSVAGEKEIKYEVDVLVPAIEGGDE 1740  
 QY 183 -----LL-----VSGS-----G-----D-----S--191  
 DB 1741 TSYFIVMNNLELDCHVTGSPPTIMWLKDGOLIDERDGFKILLNGRKLVIQAQVSN 1800  
 QY 192 -----TA-----R-----I--W 196  
 DB 1801 GLYRCMAANTAGDKKKEFEVTVHPPTIKSSGLSERVVVKYKVALQCIANGIPNSITW 1860  
 QY 197 -----NL-----S-----E-----NS--T-----S 204



QY 467 -----H-----I-W----- 469  
Db 4079 PPVISPPLKEVYIAVDKPIITLSCADGLPPDITWHKDGRAIVESINQRVLSGSLQIAF 4138  
QY 470 -----N-----T-----Q----- 472  
Db 4139 VQPDAGHYTCMAANVAGSSSTKLTVHVPPRIIRSTEGHYTVNENSOAILPCVADGPT 4198  
QY 473 -----T-----G-----A-----L-VH 478  
Db 4199 PAINWKNDVLANLLGKYTAEPYGELILENVVLEDSGFYTCVANNAAAGDTHVSLTVH 4258  
QY 479 -----S-----Y-----R 481  
Db 4259 VLFTFTELPDGVSLNKGEQLRLSCKATGIPLPKLTWTFNNNIIPAHFDSVNGHSELVIER 4318  
QY 482 -----GT-----G-----IF-----E-----V-C----- 490  
Db 4319 VSKEDSGTYVCTAENSFGVFKATGFVVVKPPFKGDPNSNWIPLGGNAILNCEVKGDP 4378  
QY 491 -----WN-----A-----AGD----- 496  
Db 4379 TPTIQNRKGVDEISHRIQLGNSLAITYGTVNEDAGDYTCVATNEAGVVERSMSTLQ 4438  
QY 497 -----K-V-----G-----A-----SAS----- 503  
Db 4439 SPPITLEPVETVINAGKKIILNQATGEBPOPTITWSRQHSISWDRVNLNNSLYIA 4498  
QY 504 -----D-----GSV-----C-V----- 509  
Db 4499 DAQKEDISEFECVARNLMGSLVLRVPVIVQVHGFSQWSAWRACSVTCGKGIQKRSCLN 4558  
QY 510 --L-----DL--R-----K 514  
Db 4559 QPLPANGKPCQGSDLERNRCONK 4582  
RESULT 14  
ADJ83137  
ID ADJ83137 standard; protein; 5636 AA.  
XX AC ADJ83137;  
XX AC  
XX AC  
DT 06-MAY-2004 (first entry)  
XX DE Human hemikentcin protein - SEQ ID 128.  
XX KW NOVX; cytostatic; antiarteriosclerotic; antidiabetic; antiasthmatic;  
KW antiallergic; antiinflammatory; respiratory; antiarthritic;  
KW dermatologic; antibacterial; cerebroprotective; vasotropic; cardiant;  
KW haemostatic; hypotensive; hepatotropic; neuroprotective; anorectic;  
KW nootropic; antitumor; muscular; immunosuppressive; gynaecological;  
KW antipsoriatic; endocrine; ophthalmological; osteopathic;  
KW antiparkinsonian; anticonvulsant; tranquiliser; analgesic; nephrotropic;  
KW antinfertility; antilipemic; cardiomyopathy; atherosclerosis; diabetes;  
KW cell signal processing; metabolic pathway; asthma; allergy; emphysema;  
KW autoimmune; graft-versus-host; arthritis; cancer; stroke; haemophilia;  
KW obesity; Alzheimer's; pain; chromosome mapping; tissue typing; human;  
hemikentcin.  
XX OS Homo sapiens.  
XX OS  
PN US2003170630-A1.  
XX PN  
XX PD 11-SEP-2003.  
XX PF 21-DEC-2001; 2001US-00032189.  
XX PF 21-DEC-2000; 2000US-0257495P.  
PR 22-DEC-2000; 2000US-0258171P.  
PR 20-FEB-2001; 2001US-0269940P.  
PR 08-MAR-2001; 2001US-0274192P.

PR 22-MAR-2001; 2001US-0277826P.  
PR 23-MAR-2001; 2001US-0279840P.  
PR 11-APR-2001; 2001US-0282981P.  
PR 13-APR-2001; 2001US-0283656P.  
PR 31-JUL-2001; 2001US-0309247P.  
PR 10-AUG-2001; 2001US-0311754P.  
PR 17-AUG-2001; 2001US-0313331P.  
XX (ALSO/) ALSOBROOK J P.  
PA (TCHE/) TCHERNEV V T.  
PA (LIUX/) LIU X.  
PA (SPYT/) SPYTEK K A.  
PA (ZERH/) ZERHUSEN B D.  
PA (PATI/) PATTURAJAN M.  
PA (LEPL/) LEPLEY D M.  
PA (BURG/) BURGESS C E.  
PA (SHIM/) SHIMKETS R A.  
PA (GROS/) GROSSE W M.  
PA (SZEK/) SZEKERES E S.  
PA (VERN/) VERNET C A M.  
PA (LILL/) LI L.  
PA (CASM/) CASMAN S J.  
PA (BOLD/) BOLDOG F L.  
PA (GORM/) GORMAN L.  
PA (GANG/) GANGOLLI E A.  
PA (FERN/) FERNANDES E R.  
PA (RIGG/) RIEGER D K.  
PA (EDIN/) EDINGER S R.  
PA (GUNT/) GUNTHER E.  
PA (MILL/) MILLET I.  
PA (SCIO/) SCIORE P.  
PA (ELLE/) ELLERMAN K.  
PA (MACD/) MACDOUGALL J R.  
PA (SMIT/) SMITHSON G.  
XX  
PI Alsobrook JP, Tchernev VT, Liu X, Spytek KA, Zerhusen BD;  
PI Patturajan M, Lepley DM, Burgess CE, Shimkets RA, Grosse WM;  
PI Szekeres ES, Vernet CAM, Li L, Casman SJ, Boldog FL, Gorman L;  
PI Gangolli EA, Fernandes ER, Rieger DK, Edinger SR, Gunther E;  
PI Millet I, Sciore P, Ellerman K, Macdougall JR, Smithson G;  
XX WPI; 2003-898249/82.  
XX  
PT New NOVX polypeptides and nucleic acid molecules, useful for diagnosing,  
PT preventing or treating NOVX-associated polypeptide disorder, e.g.  
PT cardiomyopathy, atherosclerosis, diabetes, cancer, Parkinson's disease or  
PT asthma.  
XX  
PS Disclosure; SEQ ID NO 128; 263pp; English.  
XX  
CC The invention relates to a novel isolated NOVX polypeptide. The  
CC polypeptide demonstrates cytostatic, antiarteriosclerotic, antidiabetic,  
CC antiasthmatic, antiallergic, antiinflammatory, respiratory,  
CC antiarthritic, dermatologic, antibacterial, cerebroprotective,  
CC vasotropic, cardiant, haemostatic, hypotensive, hepatotropic,  
CC neuroprotective, anorectic, nootropic, antitumor, muscular,  
CC immunosuppressive, gynaecological, antipsoriatic, endocrine,  
CC ophthalmological, osteopathic, antiparkinsonian, anticonvulsant,  
CC tranquiliser, analgesic, nephrotropic, antinfertility and antilipemic  
CC activities. The NOVX polypeptide, nucleic acid or antibody of the  
CC invention may be useful for treating or preventing a NOVX-associated  
CC disorder, such as cardiomyopathy, atherosclerosis, diabetes or a disorder  
CC related to cell signal processing and metabolic pathway modulation.  
CC Furthermore, the NOVX polypeptides may be useful for diagnosing, treating  
CC or preventing diseases such as asthma, allergies, emphysema, autoimmune  
CC disease, graft-versus-host disease, arthritis, cancer, stroke,  
CC haemophilia, obesity, Alzheimer's disease and pain. The nucleic acids may  
CC be used as hybridisation probes, in chromosome mapping, tissue typing, of  
CC preventive medicine or pharmacogenomics. The current sequence is that of  
CC a protein of the invention which is related to human NOVX protein.  
XX  
SQ Sequence 5636 AA;



Query Match 71.6%; Score 2627.8; DB 7; Length 5636;  
Best Local Similarity 9.6%; Pred. No. 3.9e-31;  
Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;

QY 1 M-S-----I-----SS--D-----EV-----8  
DB 1 MISWEVHTVFLFALLYSSLAQASQSEIRABEFPAGSTLAFVFDVTGSMYDDLQVI 60  
QY 9 -----NF-LV-----12  
DB 61 EGASKILETLKPKPLNFALVPFHPDPEIGPVTTITDPKFKQYBELRELYVOGGDCPE 120  
QY 13 -----Y-----R-----Y-----LQ-----17  
DB 121 MSIGAIKALEISLPGSFYVFTDASKDYRLTHEVLQLIQKQSQVFWLJGDCDRTH 180  
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26  
DB 181 IGYKVEEIASTSGQVFLDKKQVNEVLKWVEAVQAKVHLLSTHLEQAVNTWRIFP 240  
QY 27 -----T-----FG-----I-----E-----31  
DB 241 DPSLKEVTVLSGSPMIEIRNPLKLIKKGFGHLLNIHNSAKVUNVKEPEAGMTVK 300  
QY 32 ---S--H-----I-----S--Q-----36  
DB 301 TSSSGRHSVRITGLSTIDPRAGFRKPTLDFKTVSRPVGIGIPTYVLLNTSGISTPARID 360  
QY 37 -----S-----NI-----N-----G-----41  
DB 361 LLELLSIGSSSLKTIPIVKYPPHKKPYGIWNISDFVPPNEAFFLKVTGYDKDYLFORVSS 420  
QY 42 -----A-----L-----VP-----P-----A 47  
DB 421 VSPSSIVDPAPKVTMEKTPGYVYLOQIPCSVDLSLLPFTLSFRNGVTLGVQYKESA 480  
QY 48 -----A-----L-----I-----S-----IIQ-----KG-----56  
DB 481 SVSLDIKVTLSDEGFVECIASVAGTGRAQTFDFVSEPPVQVNNVTVPGERAVLT 540  
QY 57 -----L--Q-----P-----Y--V--E-----61  
DB 541 CLIIISAVDYNLTQRNDRVRLAEPARITLANLSLELKSVPKFNDAAGEYCHMVSSEGGSS 600  
QY 62 -A-----EVS-----NE-----68  
DB 601 AASVFLTVQEPKVTMPKMQSTGGSEVSMCSATGPKPKIAWTVMDFIVGSHRYRM 660  
QY 69 --DGLTF-----D-----G-----R-----P-----I--78  
DB 661 TSDGTLFIKNAAPKDAIGYCLASNSAGTDKQNSTLRYIEAPKLMVYQSELLVALGDIIV 720  
QY 79 -E--S-----L--S--LI-----DA-----V-----87  
DB 721 MECKTSGIPPPQWKFGDLEAPSTFLIIDLPLGLLKIQETQDLDAAGDYTCVAINAAGR 780  
QY 88 -----M-----P-DV-----VQ-----T-----94  
DB 781 ATGKITLVGSPPVFTQEPADVSMEIGSNVTLPYVQGYPEPTIKWRRLDNMPIFSRPFS 840  
QY 95 -----R-----95  
DB 841 VSSISQRTGALFILNMAWDKTYICEAENQFKIQSETTVTTVTGLVAPLIGISPVAN 900  
QY 96 --QQ-----A-----Y--R-D-KL-----AQ-----105  
DB 901 VIEGQQLTFLPCTLLAGNPIPERWIKNSAMLLQNPYITVRSGLSHIERVQLQDGGYEYC 960  
QY 106 -----OH-----A-----108  
DB 961 VASNAGTNNKTSVVVHVLPITIQHQQILSTIEGIPVTLPCKASGNPKPSVIMWSKKGEL 1020  
QY 109 -----A--A--A-----A--A--A-----114

DB 1021 ISTSSAKFSAGADGSLYVSPGESEGYVCTATNTAGYAKRKVQLTVYVPRVFGDLRG 1080  
QY 115 -----A-----A-----116  
DB 1081 LSQDKPEISVLAGEEVTLPCEVKSLPPPIITWAKETQLISPPSPRHTFLPSSGSMKITET 1140  
QY 117 -----ATN-----Q-----Q-GSA--K-----N-----126  
DB 1141 RTSDSGMVLCVATNIAGNVQAVKLVNVHVPKIQRGPKLVQVQGVQVVDIPCAQAQGTPLP 1200  
QY 127 -----GE--N-----T--A--N--G-----133  
DB 1201 VITWSKGGSTMLVDGSHVSNPDGTLSDQATPSDAGIVTCVATNIAGTDETEITLHVQE 1260  
QY 134 -----E-----E--N-----GA-----H-----139  
DB 1261 PPTVEDLEPPYNTTPOBRVANORIEPPCPAKGTPKPTIKWLHNGRELGTREPGISILEEG 1320  
QY 140 T--IA-----N-----N--HT--DM-----P--158  
DB 1321 TLLVIASVTPDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKQVSNVSVLLNQLTNL 1380  
QY 149 -----M-----E-----150  
DB 1381 FCEVEGTPSPIIMWKDNVQVTTESSTIQTVNNGKILKLFPRATPEDAGRYSCAKAINIAGTS 1440  
QY 151 ---VD-----G-----DV--E-----I-----P--158  
DB 1441 QKYFNIDVLVPTTIIGTNFPKEVSVVLRDVALEQVKGTPTPDHFWKDGKFLGDPN 1500  
QY 159 -----SN--K-A-----V-----VL 165  
DB 1501 VELLDRGVHLKNARNDKGRYQCTVSNAAQKQADIKLTIYIPPSIKGNVTTDISVL 1560  
QY 166 -----RG-----H--168  
DB 1561 INSLIKLECKTRGLPMAITWKDQOPIMSSSOALYIDKQVLIHPRAQVSDSATYCHV 1620  
QY 169 -----E--S-----E-----VFI-----C--A-----W-----177  
DB 1621 ANVAGTAESFHVVDVVPPIEMGNLATPLNKQVIAHSLTLECNAAGNPSILTLWKDGV 1680  
QY 178 -----N-----P-V--SD--182  
DB 1681 PVKANDNFRIEAGKKLEIMSAQEDRGQYICVATSVAGEKEIKYEVDVLVPPAISGGDE 1740  
QY 183 -----LL-----VSGS-----D-----S--191  
DB 1741 TSVFIVMNNLLELDCHVTGSPPTIMLWKDGLIDERDGFKILLNKRKLVIQAQVSN 1800  
QY 192 -----TA-----R-----I-W 196  
DB 1801 GLYRCMAANTAGDHKKEFEVTVHVPPTIKSSGLSERVWVYKYPVALQCIANGIPNSITW 1860  
QY 197 -----NL--S-----E-----NS--T-----S 204  
DB 1861 LKDDQVNTAQNLKIQSSGRVLQIAKTLLELAGRYTCVATNAAGETQOHIQHVHEPPS 1920  
QY 205 --G-----PTQL-----V-----L-R-----212  
DB 1921 LEDAGKMLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLLSGTSMTFLNRGQIIDIES 1980  
QY 213 -----H-----C-----I--215  
DB 1981 AQISDAGIYKCAINSAGATELFYSLQVHVAPSIGSNMNVAVVNNVNRLECEARGIPA 2040  
QY 216 -----R-----E--G-----G--Q--D--221  
DB 2041 PSLTWLKDQSPVSSPNSGLQVLSGGRILALTSTQISDTGRTCYAVNAAGEKQRDIDLVR 2100  
QY 222 -VPSN-----KD-----VT--S-L--231

Db 2101 YVPPNIMGEQNVSVLISQAVELLCSQSDAIPPPPTLTLWKDGHPLLLKKPGLSISNRSVLK 2160  
Qy 232 -----D-----W--N--S-----EG---TLL-- 240  
Db 2161 IEDAQVODTGRYTCEATNAGKTEKKNYNWPPNIGGSDTLTQLTWIEGNLSLCE 2220  
Qy 241 AT-----GS--Y--D-----G-----F-----A- 249  
Db 2221 SSGIPPPNLIWKKGSPVLTDMSGRVILSGRQLQISIAEKSDAALYCVASNVAGTAK 2280  
Qy 250 -----R-----I-----WTKDG----- 256  
Db 2281 KEYNLQVIRPTTNSGSHPEIIVTEKGSISLECEVOGIPPPPTVTKMGDGHPLIKAKGV 2340  
Qy 257 -----NL-AS-----T-----L-----GOH----- 265  
Db 2341 EILDEGHILQKNIHVSDTGRYVCVAVNAGMTDKKYDLVSHAPPSIIGNHRSPENISVV 2400  
Qy 266 -K-----G-PI-----FALK--W----- 274  
Db 2401 EKNSVLTCEASGIPLPSTTWF--KOGFVLSNSVRILSGRMLRMQTTMEDAGQYTC 2458  
Qy 275 -----N-----K-K-----GN----- 279  
Db 2459 VVRNAAGEERKIFGLSVLPPHIVGENTLEDVKKEKQSVTLTCEVTGNVPPEITWHKDG 2518  
Qy 280 -----F--I-----L-----S---AGV- 286  
Db 2519 QPLQDEBAHHIISGGFLOITNVQVHTGRYTCLASSPAGHKSFSNLVFSVPTIAGVG 2578  
Qy 287 -D--K--T-----T--I-----I-W----- 293  
Db 2579 SDGNPEDVTILNPSLVCESYPPATTIWFKDGTPLESNNRILPGRTLOILNAQ 2638  
Qy 294 -D-----A-----H-----TG-----E 299  
Db 2639 EDNAGRYCVATNEAGEMIKHYEVKVIPIINKGDLWGFLSPKEVKIKVNTLTLECE 2698  
Qy 300 A-----K-----Q-----Q----- 303  
Db 2699 AYAPSASLWYKDGQPLKDDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758  
Qy 304 F-----P--F-----P-----LDV-----D 315  
Db 2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPIIEGANS 2878  
Qy 316 -----WQ-----SN----- 319  
Db 2879 LPBEVTVNKSALIECLSSGSPAPRNSWQOGQPLEDDHKKFLSNGRILQILNTQITD 2938  
Qy 320 -----NT-----FAS--C-----S 326  
Db 2939 IGRYVCAENTAGSAKYNLNVHVPSPVIGPKSENLTIVVNNFISLTCEVSGFPDPLS 2998  
Qy 327 -----T-----D-----MCI----- 331  
Db 2999 WLKNXQPIKLTNTLIVPGGRTLQIIRAKVSDGGEVTCIAINXAGESKKFSLTVVYVPS 3058  
Qy 332 -----H-----V--C-----K-----L--GQ--- 338  
Db 3059 IKHDSLSLVNVREGTSVLECESNAVPPPVITWYKGRMITESTHVEILADGQMLHI 3118  
Qy 339 -----D-----R-----P-IX----- 343  
Db 3119 KKAESVDTGOYVCRAINAGDDKNFHLNVVPPSIEGPEREVIVETISNPVTLTCDATG 3178  
Qy 344 -----T-----F-----Q--G-----H-----T----- 349  
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGSKLOIARSQHSDSGNYTCIASNMEGKAQKY 3238

Qy 350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
Db 3239 FLISQVPPSVAGABIPSDVSVLGENVELVANGIPTPLIOWLKDGPFIASGETERIV 3298  
Qy 361 -----G--NLL-A--S-----C--S-----D-----D-----MTL-- 373  
Db 3299 SANGSTLNIYGALTSTGKYTCVATNPAGEEDRIFNLNVVYVPTIRGNKDEAKLMTYVD 3358  
Qy 374 -----K-----I-W-----SM----- 378  
Db 3359 TSINISCRXTGTPPPQINMLKGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVVTCVASN 3418  
Qy 379 -----K--Q-----DN-----CV----- 384  
Db 3419 RAGVDNKHYNLQVAFAPPNMDNSMGTEITVLKGSSTSMACITDGTAPSMAWLRDGOPLG 3478  
Qy 385 -----H-----D--L-----QA-----H-N-----KE 393  
Db 3479 LDAHLLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFLKVLPPHINGSEHEE 3538  
Qy 394 I-----YT-----I--K--W----- 399  
Db 3539 ISVINNPLELTCIASGIPAPKMTWMDGRPLPOTQOVOTLGGGEVLRISTAQVEDTGRY 3598  
Qy 400 -----SPTG-----P--GT-----N----- 407  
Db 3599 TCLASSPAGDDDKYELVRVHVHPENIAGTDEPRDITVLRNRQVLTCKSDAVPPPVITWLR 3658  
Qy 408 N-----P-----N-----ANLM-LAS--A--S--F----- 420  
Db 3659 NGERLOATPRVILSGRYLOINNADLGDNTANYTCVASNIAGTTREFLTIVNVPNIKG 3718  
Qy 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433  
Db 3719 GPQSLVILLNKSTVLESCIAEGVPTPRITWRKQDGLVAGNHARYSILENGFLHIOASHVTD 3778  
Qy 434 -----C-----I--H-----TL-----T----- 439  
Db 3779 TGRYLCMATNAGTDRRRIDLVHVPSPSIAPGTNTMTVINVQTTLACEATGPKPSINW 3838  
Qy 440 -K--H-----Q-E-----P--V-----Y----- 446  
Db 3839 RKNHLLNVQDNQNSYRLLSSGLSVIISPSVDDTATYECTVINGAGDDKRTVDLTQVPP 3898  
Qy 447 SVA-----P--S-----P--D----- 453  
Db 3899 STADEPTDFLVTKHAPAVITCTASGVPPFSIHWTKNIGIRLLPRGDGYRILSSGAIBLAT 3958  
Qy 454 -----GRY-----L-----ASG--S-- 461  
Db 3959 QLNHAGRYTCVARNAAGSAHRVTLHVHPPVIOQOPSELHVILNNPILLPCEATGTPSP 4018  
Qy 462 F-----D-----KCV----- 466  
Db 4019 FITWQKEGINVNTSGRHAVALPSGGLOISRAVREDAGTMCVAQNPAGTALGKILNVQV 4078  
Qy 467 -----H-----I-W----- 469  
Db 4079 PPVISPHKEYIAVDKPIITLSCADGLPPPDITWHKDGRAIVESIRQVRLSSGSLQIAF 4138  
Qy 470 -----N-----T-----Q----- 472  
Db 4139 VOPGDAGHYTCMAANVAGSSSTSTKLTUVHPPRIRSTEGHYTVNENSQAILPCVADGPT 4198  
Qy 473 -----T-----G-----A-----L-VH 478  
Db 4199 PAINKKONVLLANLIGKYTABPYGELILENVLEDGSGFYTCVANNAAAGETHVSLTVH 4258  
Qy 479 -----S-----Y-----R 481  
Db 4259 VLPTFTPELFDVSLNKGEOLRLSCKATGIPKLWTFNNNIIPAHFDSVNGHSELVIER 4318

QY 482 -----GT-----G-----G-----IF-----E-----V-C----- 490  
Db 4319 VSKEDSGTVCTAENSFGVKATGFFVVKPPFKGDPGPNWIEPLGGNAILNCEVKGDP 4378  
QY 491 -----WN-----A-----AGD----- 496  
Db 4379 TPTIQNRKGVDEIEISHRIQLNGSLAIYGTWNEDAGDYTCVATNEAGVVERMSLTLO 4438  
QY 497 -----K-V-----G-----A-----SAS----- 503  
Db 4439 SPTIITLEPVEVINAGGKIILNCQATGEPQPTITWSRQGHSSISWDRVNVLSNNSLYIA 4498  
QY 504 -----D-----GSV-----C-V----- 509  
Db 4499 DAQKEDTSEPCVARNLMGSLVRVPVIVQVHGGFSQWSAWRACSVTCGKGIQKRSRLCN 4558  
QY 510 --L-----DL--R-----K 514  
Db 4559 QPLPANGKPCQGSDBEMRNCONK 4582

RESULT 15

ADK60205  
ID ADK60205 standard; protein; 5636 AA.

XX AC ADK60205;  
XX DT 06-MAY-2004 (first entry)

Angiogenesis differentially expressed protein GS-P29.

XX vasotropic; antirheumatic; antiarthritic; hypotensive; antiangiogenic;  
XX antiinflammatory; cardiant; angiogenesis inhibitor; gene therapy;  
XX angiogenesis; endothelial cell; diagnosis; tumor vascularization;  
XX retinopathy; rheumatoid arthritis; Crohn's disease; atherosclerosis;  
XX ovary hyperstimulation; psoriasis; endometriosis; restenosis;  
XX angioplasty; cicatrization; peripheral vascular disease; hypertension;  
XX vascular inflammation; Raynaud's disease; aneurism; thrombophlebitis;  
XX ischemia; angina; myocardial infarction; chronic heart disease;  
XX cardiac congestion; macular degeneration; osteoporosis.

XX Homo sapiens.

XX FR2836687-A1.

XX 05-SEP-2003.

XX 11-APR-2002; 2002FR-00004546.

XX 04-MAR-2002; 2002FR-00002717.

XX (GENE-) GENE SIGNAL.

XX (ALMA/) AL MAHMOOD S.

XX Colin S, Schneider C, Al Mahmood S;

XX WPI; 2004-013912/02.

XX N-PSDB; ADK60455.

XX Compositions for diagnosing, prognosing and treating angiogenic disorders  
XX including tumor vascularization and heart disease, comprise nucleic acid  
XX or polypeptide differentially expressed in angiogenesis.

XX Claim 7; SEQ ID NO 81; 424pp; French.

XX The invention relates to a novel pharmaceutical composition active on  
XX angiogenesis comprising an endothelial cell nucleic acid whose expression  
XX is induced by an angiogenic factor and inhibited by an angiostatic agent  
XX or its complement or fragment, a polypeptide sequence encoded by the  
XX nucleic acid or its fragment, a molecule capable of inhibiting expression  
XX of the nucleic acid or a molecule which binds to the polypeptide  
XX sequence. The invention is used to diagnose, prognose or treat an  
XX angiogenic disorder in a mammal, particularly a human. The disorder is

CC particularly tumor vascularization, a retinopathy, rheumatoid arthritis,  
CC Crohn's disease, atherosclerosis, ovary hyperstimulation, psoriasis,  
CC endometriosis associated with neovascularization, restenosis due to  
CC angioplasty, overproduction of tissue due to cicatrization, a peripheral  
CC vascular disease, hypertension, vascular inflammation, Raynaud disease,  
CC aneurism, arterial restenosis, thrombophlebitis, ischemia, angina,  
CC myocardial infarction, chronic heart disease, cardiac congestion or  
CC macular degeneration due to age or osteoporosis. This sequence  
CC corresponds to a protein encoded by a differentially expressed DNA used  
CC in the composition of the invention.

XX Sequence 5636 AA;

Query Match 71.6%; Score 2627.8; DB 8; Length 5636;  
Best Local Similarity 9.6%; Pred. No. 3.9e-31;  
Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;

QY 1 M-S-----I-----SS--D-----EV----- 8  
Db 1 MISWEVVHTVFLFALLYSLAQDASPSQSEIRABEPPEGASTLAFVFDVTGSMYDDLQVVI 60  
QY 9 -----NF-IV----- 12  
Db 61 EGASKILETSLKRPKRPFLNFALVPHDPEIGPVITTDPKKQYELRELYVQGGDCPE 120  
QY 13 -----Y-----R-----Y-----LQ----- 17  
Db 121 MSIGAIAKIALEISLPGSFYIVFTDARKSKDYRLTHEVLQIQKQSQVFLVTGDCDRTH 180  
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26  
Db 181 IGKVVYEETASTSGQVPHLDKKQVNEVLKVVVEAQAQSKVHLLSTDHLEQAVNTWRIPF 240  
QY 27 -----T-----FG-----I-----E----- 31  
Db 241 DPSLKEVTVLSGSPSPMIBIRNPLGKLIKKGFGHLLHINSAKVNVNKEPEAGMWTVK 300  
QY 32 ---S--H-----I-----S-----Q----- 36  
Db 301 TSSSGRHVRITGLSTIDFRAGFSRKPXTLDFKKTVSRRPVQGIPTVYLLNTSGISTPARID 360  
QY 37 -----S-----N-----N-----G----- 41  
Db 361 LLELLISGSSSLKTI PVKYVPHRKPYGIWNISDFVPPNEAFFLKVTGYDKDDYLFQVSS 420  
QY 42 -----A-----L-----VP-----P-----A 47  
Db 421 VSFSSIVDPAPKVTWPEKTPGYLQPGQIPCSVDSLLPFTLSFVRNGVTGLGVQYLKESA 480  
QY 48 -----A--L-----I-----S-----IIQ-----KG----- 56  
Db 481 SVSLDIKVTLSDEGFYECIAVSSAGTGRAQTFFDVSEPPPIQVNNVTVTGERAVLT 540  
QY 57 -----L--Q-----Y-----V--E----- 61  
Db 541 CLIIISAVDYNLTQWRNDRVRLAEPARIRTLANLSLELSKVKENDAGEYHCWVSSEGSS 600  
QY 62 -A-----EVS-----NE----- 68  
Db 601 AASVFLTVQEPKVTVMKQNSFTGGSEVSMCSATGPKPKIAVTNDMFVGSHRMYM 660  
QY 69 --DGLTF-----D-----G-----R-----P-----I-- 78  
Db 661 TSDGTLFIKNAAPKADAGIYGLASNSAGTDKQNSTLRYTEAPKLMVMVQSELLVALGDITV 720  
QY 79 -E--S-----L--S--LI-----DA-----V----- 87  
Db 721 MECKTSGIPPPQVQWFKGDLERPSFTLLIDPLGLLKKIQETQDLQDAGDYTCVAINAGR 780  
QY 88 -----M-----P-DV-----VQ-----T----- 94  
Db 781 ATGKITLDVGSPPVFTQEPADVSMIEGNSVNTLPCYVQGVPEPTIKWRRLDNMFISRPFS 840

QY 95 -----R----- 95  
Db 841 VSSISQLRTGALFILNLWASDKTYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVAN 900  
QY 96 -----QQ-----A-----Y---R-D-KL-----AQ----- 105  
Db 901 VIEGOQLTLPCTLLAGNPIPERRWIKNSAMLLQNPYITVRSGLSHIERVOLQDGGEYTC 960  
QY 106 -----OH-----A----- 108  
Db 961 VASNAGTNNKTTSSVVHVLPTIQHQQLLSTIEGIPVTLPCKAGNPKPSVIMSKKGL 1020  
QY 109 -----A---A-A-----A---A-A----- 114  
Db 1021 ISTSSAKFSAGADGSLVYVSPGEESGEYVCTATNAGYAKRKVQLTYYVPRVFGDLRG 1080  
QY 115 -----A-----A----- 116  
Db 1081 LSQDKPVEISVLAGEEVLTPCEVKSLLPPIITWAKETQLISPFSPRHTPLPSSGSMKITET 1140  
QY 117 -----ATN-----Q-----Q-GSA-K-----N----- 126  
Db 1141 RTSDSGMYLCVATNIAGNTQAVKLVNHPVKIQRPKHLKVQGORVDIPCAQGTPLP 1200  
QY 127 -----GE---N-----T-A-N-G----- 133  
Db 1201 VITWSKGSNMLVDGHHVSNPDGTLSDIDQATPSDAGIYTCVATNAGTDETEITLHVQE 1260  
QY 134 -----E-----E---N-----GA-----H----- 139  
Db 1261 PPTVEDLEPPYNTTFQERVANORIEFPCPAKGTPKPTIKWLHNGRELTGREPGISILEEG 1320  
QY 140 T---IA-----N-----N---HT---DM----- 148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHPVVIKDEQVSNVSVLLNQLTNL 1380  
QY 149 -----M-----E----- 150  
Db 1381 FCEVEGTPSIIMWYKDNQVTESSITQTVNNGKILKLFRATPEDAGRYSCKAINAGTS 1440  
QY 151 -----VD-----G-----DV-E-----I-----P- 158  
Db 1441 QKYFNIDVLVPTIIGTNFPKEVSVNLNRDVALEQVKGTPFPDIHWFKDGKPLFLGDPN 1500  
QY 159 -----SN---K-A-----V-----V-----VL 165  
Db 1501 VELLDRGQVLHLKNARNNDKGRYQCTVSNAAQKQAKDIKLTIVIPPSIKGNVTTDISVL 1560  
QY 166 -----RG----- 168  
Db 1561 INSLIKLECKTRGLPWPATWYKDGQPIMSQALYIDKQYLHI PRAQVSDSATYTCV 1620  
QY 169 -----E-S-----E-----VFI-----C-A-----W----- 177  
Db 1621 ANVAGTAESFHVVDVYVPPMIEGNLATPLNKQVIAHSLTECNAGNPSPIITLWKDGV 1680  
QY 178 -----N----- 182  
Db 1681 PVKANDNFRIEAGKKLEINSAQEI DRGQYICVATS VAGEKEIKYEVDVLVPPAIEGGDE 1740  
QY 183 -----LL-----VSGS-----G-----D-----S----- 191  
Db 1741 TSYFIVMNNLELDCHTVTSPPPTIMWLKDGQLIDERGFKILLNGRKLVI AQVSN 1800  
QY 192 -----TA-----R----- 196  
Db 1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSSGLSERVVVYKPVALQCIANGIPNPSITW 1860  
QY 197 -----NL---S-----E-----NS---T-----S 204  
Db 1861 LKDDQPNTAQNKLKIQSSGRVLIQAKTLLEDAGRYTCVATNAGETQQHQLHVHPPS 1920  
QY 205 -----G-----PTQL-----V-----L-R----- 212

Db 1921 LEDAGMNLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLLGGSTMTFLNRQIIDIES 1980  
QY 213 -----H-----C-----I--- 215  
Db 1981 AQISDAGIYKVAINSAGATELFYSLQVHVAPFISGSNNMVAVVNNVPRLECEARIPA 2040  
QY 216 -----R-----E-G-----G-Q---D--- 221  
Db 2041 PSLTWLKDGPVSSPFSNGLOVLSGGRI LALSTQISDTGRTYTCVAVNAAGEKORDIDLRV 2100  
QY 222 -VPSN-----KD-----VT---S-L- 231  
Db 2101 YVPPNIMBEEONVSVLISOAVELLQSDAI PPPTLTWLDKGHLLKPKGLSISENSRVLK 2160  
QY 232 -----D-----W---N---S-----EG---TLL--- 240  
Db 2161 IEQAQVQDGTGRYTCEATNVAGTEKKNVNNVWPPNIGGSDELDTOLTWIEGNLISLLCE 2220  
QY 241 AT-----GS--Y-D-----G-----F-----A- 249  
Db 2221 SSGIPPNLIWKKGSPVLTDMSGRVRIILSGGRLOISAEKSDAALYSCVASNVAGTAK 2280  
QY 250 -----R-----I-----WTKDG----- 256  
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSI SLECEVOGIPPTVTWMDKGHPLIKRAKV 2340  
QY 257 -----NL-AS-----T---L-----COH----- 265  
Db 2341 EILDEGHILQKNIHVSDDTGRVYCVAVNVAGTMDKKYDLSVHAPPSIIGNHRSPENISVV 2400  
QY 266 -K-----G-PI-----FALK-W----- 274  
Db 2401 EKNSVLTCEASGIPLPSTWPF--KDCWPSVLSNSVRILSGGRMLRLMQTTMEDAQYTC 2458  
QY 275 -----N-----K-K-----GN----- 279  
Db 2459 VVRNAAEERKIFGLSVLPVPHVGTLEDVKVKEQSVTLTCEVTGNPVPBITWHKDG 2518  
QY 280 -----F-I-----L-----S---AGV- 286  
Db 2519 QPLQDEAHHIISSGGRFLOITNVQVPHTRGYTCCLASSPAGHKRSRPSLNVFVSPTIAGV 2578  
QY 287 -D---K-T---T-I-----I-W----- 293  
Db 2579 SGNPEDVTILNSPTSLVCEAYSYPATITWFKDGTPLESNRNIIRILPGRTLIILNAQ 2638  
QY 294 -D-----A---H-----TG-----E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHVEKVYIPIINKGDLWGLSPKPKVKNVNTLLECE 2698  
QY 300 A-----K-----Q---Q----- 303  
Db 2699 AYAI PSASISWYKDGOPKLSDDHVNIAANGHTLQI KEAQISDGTGRYTCVASNAGEBELD 2758  
QY 304 F-----P-F-----A-----LDV-----D 315  
Db 2759 FDNVIOVPSFQKLWEIGNMLDTRNGEAKDVIINNPISLYCETNAAPPTLTWYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315  
Db 2819 LTSDDKVLILPGGRVLQIPRAKVEDAGRVTCAVNEAGEDSLQYDVRVLVPPII EGANS 2878  
QY 316 -----WQ-----SN----- 319  
Db 2879 LPBEVTVLNKSALIECLSSGSPAPRNSWQKQPLLEDHKKFLSNGRILQILNTQITD 2938  
QY 320 -----NT-----FAS-C-----S 326  
Db 2939 IGRYVCVAENTAGS AKYFNLVHPVPSVIGPKSENITVVVNNFISITCEVSGPPPDLS 2998  
QY 327 -----T-----D---MCI----- 331

Db 2999 WLKXQPIKLTNTLIVPGRTLIIRAKVSDGCVTCIAXNAGESKKFSLTVVPPS 3058  
QY 332 ---H---V---C---K---L---GO--- 338  
Db 3059 IKHDSLSLVNREGTSVLECESENAVPPVITWYKGRMITESTHVEILADQOMLHI 3118  
QY 339 ---D---R---P---IK--- 343  
Db 3119 KKAESDGTQYVCRAINVAGRDDKNFHLNVVPPSIEGPPEVIVETISNPVLTCDATG 3178  
QY 344 ---T---P---Q---G---H---T--- 349  
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGSKJQIARSOHSDSGNYTCIASNMEGKAQKY 3238  
QY 350 ---N---E---V---NA---IKW---D---P---T--- 360  
Db 3239 FLSIQVPPSVAGAEIPSDSVLGENVELVCNANGIPTPLIQLKDKGPASGETERIRV 3298  
QY 361 ---G---NLL---A---S---C---S---D---D---MTL--- 373  
Db 3299 SANGSTLNIYGALTSDTGKYTCVATNPAGEBDRIFNLNVVYVTPITIRGNKDEAEKLTMYD 3358  
QY 374 ---K---I---W---SM--- 378  
Db 3359 TSINIECRXTGTPPPQINMLKNGLPFLSSHIRLAAGQVIRIVRAQVSDVAVYTCASN 3418  
QY 379 ---K---Q---DN---CV--- 384  
Db 3419 RAGVDKNHYNLQVAPPNMDSNGTBEITVLKGSSTSMACITDGTAPSPMAWLRDQPLG 3478  
QY 385 ---H---D---L---QA---H---KE 393  
Db 3479 LDAHLTVSTHGMVLQLKAEFTDSGKYTCIASNEAGEVSKHFLKVLPEPHNGSEHEE 3538  
QY 394 I---YT---I---K---W--- 399  
Db 3539 ISVIVNPLELTCIASGIPAPKMTWMMKGRPLPQDQVOTLGGGEVLRISTAQVEDTGRY 3598  
QY 400 ---SPTG---P---GT---N--- 407  
Db 3599 TCLASSPAGDDKXELVVRVHPVNIAGTDEPRDITVLRNQVTLCKSDAIVPPPVITWLR 3658  
QY 408 N---P---N---ANLM---LAS---A---S---F--- 420  
Db 3659 NGERLQATPRVRLSGERYLIQINNADLGDYNTCVASNIAGKTITREFILTVNVPPNIG 3718  
QY 421 ---DSTV---RL---W---D---V---DR---G---I--- 433  
Db 3719 GPQSLVILLNKSTVLECIAEGVPTPRITWRKQGAVLAGNHARYSILENGFLHIQSAHVD 3778  
QY 434 ---C---I---H---TL---T--- 439  
Db 3779 TGRYLCMATNAAGTDRRIDLQVHVPSPSAPGTNTMTVIVNQVTLACEATGIPKPSINW 3838  
QY 440 -K---H---Q---E---P---V---Y--- 446  
Db 3839 RKNGHLLNVQDNQNSYRLLSSGSLVIPSVDYATYECTVTNGAGDKRTVDLTVOVPP 3898  
QY 447 SVA---P---S---P---D--- 453  
Db 3899 STADEPTDFLVTKHAPAVITCTASGVFPFPIHWTKNKIRLLPRGDGYRILSSGAIBLAT 3958  
QY 454 ---CRY--- 461  
Db 3959 QLNHAGRYTCVARNAAGSAHRVTLHVHPPVITQPOPSLHVILNPNILLPCEATGTPSP 4018  
QY 462 F--- 466  
Db 4019 FITWQEGINVTSGRHAHVLPSSGGLQISRAVREDAGTYMCVAQNAGTALGKILNVQV 4078  
QY 467 ---H---I---W--- 469  
Db 4079 PPVISPHLKEYVIAVDKPTLSCEADGLPPPDITWHDGKRAIVESIRQVRLSSGSLQIAF 4138

QY 470 ---N---T---Q--- 472  
Db 4139 VQPEDAGHYTCMAANVAGSSSTSKLTIVHVPPIRSTEGHYTVNENSOAILPCVADGIPT 4198  
QY 473 ---T---G---A---L---VH 478  
Db 4199 PAINWKKDNVLLANLLGKYTAEPYGBELILENVVLEDSGFYTCVANNAAAGDTHTVSLTVH 4258  
QY 479 ---S---Y---R 481  
Db 4259 VLPTFTELPGDVSINKGBQLRLSCKATGIPKLWTWTFNNIIPAHFDSVNGHSELVIER 4318  
QY 482 ---GT---G---G---IF---E---V---C--- 490  
Db 4319 VSKEDSGTYVCTAENSVMGFVKATGFVYVKEPPVFKGDYPSNWIEPLGGNAILNCEVKGDP 4378  
QY 491 ---WN---A---AGD--- 496  
Db 4379 TPTIQWNRKGVDIIEISHIRQLNGSLAIYGTVNEDAGDYTCVATNEAGVVERSMSTLQ 4438  
QY 497 ---K---V---G---A---SAS--- 503  
Db 4439 SPPIITLBPVETVINAGKIIILNCQATGEPQPTITWSROGHSISWDDRVNVLNNSLYIA 4498  
QY 504 ---D---GSV---C---V--- 509  
Db 4499 DAQKEDTSBPECVARNLMGSLVLRVPVIVQVHGGFQSWANRACSVTCGKGIQKRSLCN 4558  
QY 510 ---L---DL---R---K 514  
Db 4559 QPLPANGKPKCGSDLEMRNCQK 4582

Search completed: January 3, 2005, 15:20:14  
Job time : 111.333 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:10:17 ; Search time 23.6667 Seconds  
(without alignments)  
1440.315 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 3669

Sequence: 1 MSISSDEVNFLVRYLQESG.....GDKVCASGDSVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued Patents AA:\*

1: /cgn2\_6/prodata/1/iaa/5A\_COMB.pep.\*

2: /cgn2\_6/prodata/1/iaa/5B\_COMB.pep.\*

3: /cgn2\_6/prodata/1/iaa/6A\_COMB.pep.\*

4: /cgn2\_6/prodata/1/iaa/6B\_COMB.pep.\*

5: /cgn2\_6/prodata/1/iaa/PCTRUS\_COMB.pep.\*

6: /cgn2\_6/prodata/1/iaa/backfilese1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2586.1	70.5	4544	1	US-08-469-486-52
2	2586.1	70.5	4544	2	US-08-469-486-52
3	2583.8	70.4	3461	3	US-09-334-220-2
4	2577.9	70.3	4654	3	US-08-476-515A-84
5	2577.8	70.3	4655	3	US-08-452-877-84
6	2577.8	70.3	4655	3	US-08-452-877-88
7	2577.8	70.3	4655	3	US-08-452-877-90
8	2576.8	70.2	4655	3	US-08-452-877-86
9	2560.4	69.8	5405	3	US-08-718-388-9
10	2557.9	69.7	3460	3	US-09-334-220-1
11	2550.8	69.5	15281	2	US-08-471-119A-2
12	2548.4	69.5	4302	3	US-09-052-469-8
13	2548.4	69.5	4302	4	US-08-422-582-8
14	2548.4	69.5	4302	4	US-09-052-262-8
15	2548.4	69.5	4339	3	US-09-052-469-6
16	2548.4	69.5	4339	4	US-08-422-582-6
17	2548.4	69.5	4339	4	US-09-052-262-6
18	2547.1	69.4	3571	4	US-09-911-842A-2
19	2546.3	69.4	4303	2	US-08-460-751-2
20	2546.3	69.4	4303	4	US-09-479-467A-2
21	2545.4	69.4	4302	3	US-08-658-136-5
22	2541.3	69.3	4861	4	US-09-919-497-70
23	2508.2	68.4	3594	4	US-09-911-842A-4
24	2490.2	67.9	3959	2	US-08-970-269A-30
25	2490.2	67.9	3959	3	US-09-407-562-30
26	2486.4	67.8	7257	3	US-09-335-409-5
27	2486.4	67.8	7257	3	US-09-568-102-5

28	2486.4	67.8	7257	3	US-09-567-969-5	Sequence 5, Appli
29	2486.4	67.8	7257	3	US-09-568-480-5	Sequence 5, Appli
30	2486.4	67.8	7257	3	US-09-568-486-5	Sequence 5, Appli
31	2486.4	67.8	7257	3	US-09-568-472-5	Sequence 5, Appli
32	2486.4	67.8	7257	3	US-09-567-899-5	Sequence 5, Appli
33	2485.4	67.7	3727	2	US-08-222-617A-27	Sequence 27, Appli
34	2485.4	67.7	3778	2	US-08-222-617A-2	Sequence 2, Appli
35	2478.9	67.6	4866	4	US-09-424-783-2	Sequence 2, Appli
36	2477.4	67.5	3666	2	US-08-222-617A-12	Sequence 12, Appli
37	2463.8	67.2	3801	2	US-08-822-445-10	Sequence 10, Appli
38	2463.8	67.2	3801	3	US-09-396-540-10	Sequence 10, Appli
39	2458.7	67.0	4928	3	US-09-036-987A-5	Sequence 5, Appli
40	2458.7	67.0	4928	3	US-09-370-700-5	Sequence 5, Appli
41	2458.7	67.0	4928	4	US-09-603-207-5	Sequence 2, Appli
42	2457.1	67.0	3898	3	US-08-750-717-2	Sequence 2, Appli
43	2454.7	66.9	5588	3	US-09-036-987A-6	Sequence 6, Appli
44	2454.7	66.9	5588	3	US-09-370-700-6	Sequence 6, Appli
45	2454.7	66.9	5588	4	US-09-603-207-6	Sequence 6, Appli

## ALIGNMENTS

### RESULT 1

US-08-469-486-52

; Sequence 52, Application US/08469486

; Patent No. 5739281

; GENERAL INFORMATION:

; APPLICANT: Thoeergersen, Hans Christian

; APPLICANT: Holtet, Thor Las

; APPLICANT: Etzerodt, Michael

; TITLE OF INVENTION: Improved method for the refolding of

; TITLE OF INVENTION: proteins

; NUMBER OF SEQUENCES: 58

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version

; SOFTWARE: #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/469,486

; FILING DATE:

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/192,060

; FILING DATE: February 4, 1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Paul T. Clark

; REGISTRATION NUMBER: 30,162

; REFERENCE/DOCKET NUMBER: 06363/002001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617 542 5070

; TELEFAX: 617 542 8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 52:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 4544 amino acids

; TYPE: amino acid

; STRANDEDNESS:

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-469-486-52

Query Match 70.5%; Score 2586.1; DB 1; Length 4544;

Best Local Similarity 10.0%; Pred. No. 2.4e-32;

Matches 432; Conservative 65; Mismatches 8; Indels 3809; Gaps 387;

```
Qy 1 MS-----I-----SSDE-----V-----N-F-L-----11
Db 88 MSRLCNGVQDCMDGSDGPHCRELQGNCSRLGCOHCHVPTLDGTCYCNSSFQADGKT 147
Qy 12 -----VY-----R-----Y-LQ-----E-----18
Db 148 CKDFDECSVGTCSQLCTNTDGSFICGCVGYLLQPNRSCAKNEVPDRPPVLLIANSQ 207
Qy 19 -----SG-----FS-----H-SAF-T-----F 28
Db 208 NILATYLSGAQVSTITPTSTQTTAMDFSYANETVCWVHVGDSAAQQLCARMPGLKGF 267
Qy 29 GI-ESH--IS-----Q-----S-N-----I-----NG-----41
Db 268 -VDE-HTINISLSLHVQEMAIDMLTGNFYFVDDIDRIFVCRNRNGDTCVTLDLLELYNP 325
Qy 42 ---AL-----VP-----P-----A- 47
Db 326 KGIALDPAMGVFFTDYGOIPKVERCDMGQNRKLVDSKIVFPHGITLDLVSRLVTWAD 385
Qy 48 A---LI-----SIQKG-L-Q-Y---V-E-----A-EVS-I-----66
Db 386 AYLDYIEWDYEGKGRQTIIO-GILIEHLYGLTVFENLYATNSDNANAQKTSVIRVNR 444
Qy 67 -N-E-----D-GTL-----:-----72
Db 445 FNSTEYQVTVRVDKGGALHTYHQRRQPRVRSHACENDQYKPGGCSDCILLANSHKARTC 504
Qy 73 ---F---DG-----R-----PIE-----SL 81
Db 505 RCRSGFLSGDGSKCKPEHEFLVYKGRPGIIRGMDGAKVDPDEHMFIEIENLMNPRAL 564
Qy 82 -----S-LI-----D-----AV-M-----PD 90
Db 565 DPHAETGIYFADTTSYLIQKQIDGTERETILKGIHNVGVAVDMGMDNLYWTDGPK 624
Qy 91 -V-V-----QTR---Q---QA-----Y-----99
Db 625 KTSVARLEKAAQTRKTLIEGKTHPRAVVDPLNGMYWTDWEEDPKDSRRGRLERAMW 684
Qy 100 ---RD---K-----L---A---Q-----Q-----H 107
Db 685 DGSHRDI FVTKVLPNGLSLDIPAGRLYVDAFYDRIETILLNGTDKIVYEGPELNH 744
Qy 108 A-----A-----A-----AA-----111
Db 745 AFGLCCHGNLYFWTEYRSGSVYRLERGVGGAPPTVTLLRSERPPIFEIRMYDAQQQVGT 804
Qy 112 -----AA-----A-A---A---AT---N-----Q-----120
Db 805 NCRVNNGGSSICLATPGSRQACAEQVLDADGVTCLANPSYVPPQCPQGEFACANS 864
Qy 121 ---Q-----G-----S---A---K-----N---G-----127
Db 865 RCTQERWKCDGNDCLNDEAPALCHQHTCPSDRFKCNENRCIPNRWLCGDGNDGENSE 924
Qy 128 -E-N-T-A---N-----G-----E-----E---135
Db 925 DESNATCSARTCPPNQFSCASGRICIPISWTCDLDDCGDRSDESASCAVPTCPFLTQFTC 984
Qy 136 -NG-----A---H---T-----I-----141
Db 985 NNGRCININWRCDNDGNSDEAGCSHSCTQPKNSGRICIPHWTCDGNDGCDYS 1044
Qy 142 ---AN---N-----HTD-----M-----M---E---V---151
Db 1045 DETHANCTNQAATPPGCHTDEFQCRLDGLCIPLRWRCDGTDTCMDSSDEKSCGTHVC 1104
Qy 152 D-----G-D-----V-----E-I---PS-----NKAV-----163
Db 1105 DPSVKFGCKDSARCISKAWVCDGNDGCDNDEENDESACRPSPSHPCANNTSVCLPPDK 1164
```

```
Qy 164 V-----L-R-----G-H-----E-S-----E-----V- 172
Db 1165 LCDGNDGDSDEGLCDQCQSLNCGCSNCSVAFPEGIVCSPGLMELGPDNHTCQIQ 1224
Qy 173 -----F-I-CA---W-----N-----178
Db 1225 SYCAKHLKCSQCDQNKFSVKSCYEGVLEPPDGESCRSLDPPKPIIFSNRHEIRIDL 1284
Qy 179 -----P-----V-----S-----D-----LL-----184
Db 1285 HKGDYSVLVGLRNTIALDFHLSQSALYWTDDVEDKIYRKLLDNGALTSEFVIOYGLA 1344
Qy 185 -----V-S-----GS-----GD-----S-----T 192
Db 1345 TPGLAVDWIAGNIYVWESNLDQIEVAKLDGTLRTLLAGDIEHPRAIALDPRDGLFWT 1404
Qy 193 ---A---RI-----W-N-----197
Db 1405 DWDASLPRIEASMSGAGRTVHRETGSGGWPNGLTVDYLEKILWIDARSDAIYSARYD 1464
Qy 198 -----LS-----E-----NST-----SG-----205
Db 1465 GSGHMEVLRGHEFLSHPPFAVTLYGGEVYWTDMRTN-TLAKANKWTGHNTVQKTNTPF 1523
Qy 206 ---PT-----Q-----LV-L-R-----H---C-----214
Db 1524 DLQVHPSRQPMAPNCPCEANGGQPCSHLCLINYNRTVSCACPHLMKMLHKDNTCYEFKK 1583
Qy 215 -----I-----RE-----217
Db 1584 FLYIARQMEIRGVLDLAPYNYIISFTVPIDNVTVDAREQRYVSDVQTAIKRAF 1643
Qy 218 ---G-G-----QDVPSNK---D-V-----TS-----LD-----232
Db 1644 INGTGETVVSADLP-NAHGLAVDWVSRNLFWTSYDTNKKQINVARLDGSKFNAVQGLE 1702
Qy 233 -----W---N-S---EG---TLL---A---TG-----S-Y-----245
Db 1703 QPHGLVVHPLRGKLYWTDGDNISMANMDGNSRNTLLFSGQGPVGLAIDFPESKLYMISSG 1762
Qy 246 -----D-GF---A-R-----I-----W-----T-K-DG-----256
Db 1763 NHTINRCNLGSGLEVIDAMRSQKATALAMGDKLWADQVSEKMGTCCKADGSSVV 1822
Qy 257 ---N---LA-----S---T---L---G 263
Db 1823 LRNSTTLVHMVKYDESIQLDHKGTNPCSVNNGDCSQLCLPTSETTRSCMCTAGYSILRS 1882
Qy 264 Q-----HKG-----P-----I-F-A-----271
Db 1883 QQACEGVGSLYVHVEGIRGIPLDPNKSDALVPVSGTSLAVGIDFHAENDTIYVWDMG 1942
Qy 272 L-----K---W---N---K-----KGN-----F-I---L-----282
Db 1943 LSTISRAKEDQWREDVVTNGIGRVEGIADWIAIGNIYWTDOGDFDIEVARLNGSPRYV 2002
Qy 283 -SAGVDK-----T-----T-----I---I-----292
Db 2003 ISOGLDKPRAITVHPEKGYLFWTEWQGPRIERSRLDGTERRVVLNVNISPNGISVDYQ 2062
Qy 293 ---W-DA-----H-----296
Db 2063 DGKLYWCDARTDKIERIDLETGENREVLSNNMDMFSVSVFEDFIYWSDRTHANGSIKR 2122
Qy 297 ---TG-----298
Db 2123 GSKDNATDSVPLRTGIGVOLKDKIVFNDRQKGTNVCAVANGCQQLCLYRGRGORACAC 2182
Qy 299 -----E-A-----K-----Q-----Q-----P- 304
Db 2183 AHGMLAEDGASCREYAGYLLYSERTILKSIHLSDERNLNAPVQPFEDPEHMKNVIALAFD 2242
```



QY 305 -----P--F--H-----S--A----- 309  
Db 2243 YRAGTSPGTFNRIFFSDIHFGNIQINDDGSRRITIVENVSGVEGLAYHGRMDTLWTSY 2302  
QY 310 -----P-AL-----D-----V-D-----W-Q----- 317  
Db 2303 TTSTITRHTVDQTRPGAFERETVITMSGDDHPRAFLDECQNLMTWNEQHPISIMRAA 2362  
QY 318 -S--N-----N-----T----- 321  
Db 2363 LSGANVLTLEIKDITRPNGLAIDHRAELKLYFSDATLDKIERCEYDGSRYVILKSEPVHP 2422  
QY 322 -----P-----F-----A-----SC--S-- 326  
Db 2423 FGLAVYGEHI FWTDMVRRARANKHVGNMKNLLRVDIPQPMGIITAVANDTNSCFLSPC 2482  
QY 327 -----T-----DM-----CI----- 331  
Db 2483 RINNGSCODLCLLTHQGHVNCSCGRRILODLDLTCRAVNSSCRAQDEFECANGECINFSL 2542  
QY 332 -----H-----V-----C-K 335  
Db 2543 TCDGVPHCKDKDEKPSYNSRRCKTFRQCSNRCVSNMLWCNGADDCGSDSEIPCNK 2602  
QY 336 -----P-----Q-----LG----- 338  
Db 2603 TAGGVGEFRCDTCTIGNSSRCNQFVDCEDASDENMCSATDCSSYFRLGVKGLFQPCER 2662  
QY 339 -----D-----RP-----I-----K----- 343  
Db 2663 TSLCYAPSWVCDGANDCGDYSRDCPGVKRPRCLPNYPACPSGRCIPMSWTCKEDDCE 2722  
QY 344 -----T-----P-----Q-----G-----H----- 348  
Db 2723 HGEDETHCNKFCSEAFQECQNHRCISKQMLDCGSDDCGSDDEAHCEGKTCGPSFSFSCP 2782  
QY 349 -T-----N-----EV--N-A-I-K----- 356  
Db 2783 GTHVCVPERWLCDDKDCADGADESAAGLYNSTCDDREFMCQNFQICPKHFVCDHDD 2842  
QY 357 -----W-----D-----P--T----- 360  
Db 2843 CADGSDSEPECYPTGCPSEFRANCGRCLSSRQWECDEGNDCHQDSEAPKNPHCTSPER 2902  
QY 361 -----G-----N-----L----- 363  
Db 2903 KGNASSQFLCGRCAEALLCNGQDDCGSDSDRGCHINECLSRKLSGCSQDCEDLKIG 2962  
QY 364 -----L-----A-----S-----CS-----D----- 369  
Db 2963 FKCRCPGFLKDDGRTCADVDECSITTFPCSQRCINTHSYKCLCVGYAPRGDHPHCK 3022  
QY 370 -D-----M-----T-LK-----I-W-----SM 378  
Db 3023 AVTDEPFLIFANRYLRLNLDGSNTLLKQGLNNAVALDFYRQMIYVTDVTTQGS 3082  
QY 379 -K-----Q-----D-----N-----C----- 383  
Db 3083 IRRMLHNSGVQLHRTGLSNPDGLAVDWVGGNLYWCKGRDTEVSKLNGAYRTVLVSS 3142  
QY 384 -----V-----H----- 385  
Db 3143 GLREPRALVVDQVQGLYWTWDGHDLSLIGRIGMDGSSRSVIVDTKITWPNGLTLDVYTER 3202  
QY 386 -----D-----L-Q-----AH----- 390  
Db 3203 IYWADAREDYIEFASLDGSRHVLVSDIPHI FALTLFEDYVYVTDWETKSINRAHKTG 3262  
QY 391 -NK-----EI----- 395  
Db 3263 TNKTLTISTLRPMDLHVFAHQPOVPHALQPOVPHPCVNNNGGCSNCLLSPGGGHKACPTNFI 3322  
QY 396 -----T-----IK--W-----S-----P----- 401

Db 3323 LGSGRGTCVSNCTASQFVCKNDKCIFFMWKCDTEDDCGHSDPDPBPCBFKCRPGQFCS 3382  
QY 402 TG-----P-----G-----TN-----N----- 408  
Db 3383 TGICTNPAPICDGDNDQDNDEANCDIHVCLPSQPKCTNTNRCIRGIFRCNQDNGDG 3442  
QY 409 -----PN-----ANL--M----- 414  
Db 3443 EDERCPEVTCAPNQFQCSITKRCIPRVWVCDRNDNDVDSDEPANCTQMTGCVDFRCK 3502  
QY 415 -----L-A----- 416  
Db 3503 DSGRCIPARWKCDGDDCGDSEPKEDERTCEPYQFRCKNNRCVPGRWQCDYNDGC 3562  
QY 417 -----S-----A--S-----E----- 420  
Db 3563 DNSDBESCTPRCSSESEFCANGRCIAGRWKCDGDDHDCADGDEKCTPRCDMDQFOCKS 3622  
QY 421 -----D-S-----T-VR-----L-W----- 427  
Db 3623 GHCIPLRWCADADCMGSDSEACGTGVTCTPLDEFQCNNTLCKPLAMKCDGEDDCGD 3682  
QY 428 -D-----V-----DRGIC-I-----HTLYK 440  
Db 3683 SDENPEECARFVCPFNRPFRCKNDR-VCLWIGRQCDGTDCGSDGTDEDECEPPTAHT-T- 3739  
QY 441 H-----Q-----E-----P-----V 445  
Db 3740 HCKDKKBFCLRNQRCUSSLRNMFDDCGDSEDEDCSIDPKLTSCATNASICGDEARCV 3799  
QY 446 -----Y--S--V-----A-F----- 450  
Db 3800 RTEKAAACACRSGFHTVPGQCGDINECLRFGTCSQLCNNTKGGHLCSCARNFMTHT 3859  
QY 451 -----S-----P-----D-----GR----- 455  
Db 3860 CRAEGSEYQVLYIADNEIRSLPFGPHSAYEQAFQGDSESVRIDAMDVHVKAGRVYTNW 3919  
QY 456 -----Y-L--A--S-----G-----F----- 462  
Db 3920 HTGTISYRSLPAPPTTSNRHRRQIDRGVTHLNLISGLKMPRGIAIDWAGNVYWTDSGR 3979  
QY 463 D-----K-C-----V--H-I-----W-N-----T- 471  
Db 3980 DVEVAQMGKGNKTLISGMIDEPHAI VVDPLRGTYWSDGNHNPXKIEAAMDGTLRETL 4039  
QY 472 -Q--TGAL-V--HS--Y-----R-GT-----G----- 484  
Db 4040 VQDNQWPTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPIVAADS KRGLSHPFSIDVF 4098  
QY 485 -G-----I--F-----EV-----C-W 491  
Db 4099 EDYIYGVTVINNRFKIHKEGHSPLVNLTTGGLSHASDVVLYHQHKOPEVTNPCDRKKEW 4158  
QY 492 -----N-----A-----A----- 494  
Db 4159 LCLLSPSGPVCTCPNGKRLDNGTCVVPSPPTPPDAPRPTCNLCQFNGSGCFLNARQP 4218  
QY 495 -----GDK-----V--G--A-- 500  
Db 4219 KRCQPRYTGDKCELDCQWEHCRNGGTCAASPSGMTCTCPTGTGPKCTQOVQAGYCAN 4278  
QY 501 -----S-----ASDCS----- 506  
Db 4279 NSTCTVNOGNQPCRCPLGFLGDRQYRQCSYCEFGTCQMAADGSRQCRCTAYPEGR 4338  
QY 507 --V-----CVL--DL-----R-----K 514  
Db 4339 CEVNCSCRLEGACVNVKQSGDVTCTDGRVAPSLCTCVGHCSNGSGSCTMNSK 4392

```

US-08-469-658-52
; Sequence 52, Application US/08469658
; Patent No. 5917018
; GENERAL INFORMATION:
; APPLICANT: Th egersen, Hans Christian
; APPLICANT: Holtet, Thor Las
; APPLICANT: Etzerodt, Michael
; TITLE OF INVENTION: IMPROVED METHOD FOR THE REFOLDING OF
; TITLE OF INVENTION: PROTEINS
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version
; SOFTWARE: #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,658
; FILING DATE: June 5, 1995
; CLASSIFICATION: 530
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/192,060
; FILING DATE: February 4, 1994
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul T. Clark
; REGISTRATION NUMBER: 30,162
; REFERENCE/DOCKET NUMBER: 06363/002002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617 542 5070
; TELEFAX: 617 542 8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4544 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-469-658-52

Query Match 70.5%; Score 2586.1; DB 2; Length 4544;
Best Local Similarity 10.0%; Pred. No. 2.4e-32;
Matches 432; Conservative 65; Mismatches 8; Indels 3809; Gaps 387;

QY 1 MS-----I-----SSDB-----V-----N-P-L-----11
|||
88 MSRLCNGVQCDGSDGPHCBRLQNCRLGCGHCVPLTDGPTCYCNSSFOLQADGKT 147
|||
QY 12 -----VY-----R-----Y-IQ-----E-----18
|||
148 CKDFDECVYGTCSQLCTNTDGSFICGCEGYLLQPDNRSCKAKNEFVDRPPVLLIANSQ 207
|||
QY 19 -----SG-----FS-----H-SAF-T-----F 28
|||
208 NILATVLSGQVSTIPTSTROTMTDFSVANETVCWVHVGDSAAQTOLKCARMPGLKGF 267
|||
QY 29 GI-ESH-----IS-----Q-----S-N-----I-----NG-----41
|||
268 -VDE-HTINISLSLHVEQMAIDWLTGNTGFVDDIDRIFCVNRNGTCTVLLDLELYNP 325
|||
QY 42 ---AL-----VP-----P-----A-47
|||
326 KGIALDPMGKVFPTDYGQIPKVERCDMGQNTKLVDSKIVFPHTLDSRLVYVWAD 385
|||
QY 48 A---LI-----SIQKG-L-Q--Y---V-E-----A-EVS-I-----66
|||

```

Db 1465 GSGHMEVLRGHEFLSHPPFAVTLYGGEVYWTWRTN-TLAKANKWTGHNTVTVQRTNTQPP 1523  
QY 206 -PT- -Q- -LV-L-R- -H- -C- - 214  
Db 1524 DLOVTHPSQPMAPNPNCEANGGQPCSHLCLINYNRTVSCAPHLMLKHDNTTCYEFKK 1583  
QY 215 - - - - -I- -RE- - 217  
Db 1584 FLLYARQMEIRGVLDAPYNYIISFTVPIDNVTLVDAREQRYVSDVVRTQAIKRAF 1643  
QY 218 -G-G- -QDVPSNK- -D-V- -TS- -LD- - 232  
Db 1644 INGTGVETVSADLP-NAHGLAVDWVSRNLFMTSYDTNKKQINVARLDGSKFNAVVQGLE 1702  
QY 233 - - - - -W- -N-S- -EG- -TLL- -A- -TG- -S- -Y- - 245  
Db 1703 QPHGLVWHLRGKLYWTDGNSIMANMDSNRITLLFSGQKPVGLAIDFPESKLYWISSG 1762  
QY 246 - - - - -D- -GF- -A- -R- -I- -W- -T- -K- -DG- - 256  
Db 1763 NHTINRCNLGSGLEVIDAMRSQKATALAIMGDKLWADQVSEKMGTCRADGSGSVV 1822  
QY 257 -N- -LA- - - - -S- -T- -L- -G- 263  
Db 1823 LRNSTLVMHMKVYDESIQLDHKGTNPVCSVNNGDCSQLCLPTSETTRSCWCTAGYSLRS 1882  
QY 264 Q- - - - -HKG- - - - -P- -I- -F- -A- - 271  
Db 1883 QQACEGVGFLYSVHEGIRGIPLDPNKSDALVPVSGTSLAVGIDFHAENDTIYVDMG 1942  
QY 272 L- - - - -K- -W- -N- -K- -KGN- -F- -I- -L- - 282  
Db 1943 LSTISRAKRDQWREDVWVNGIRVEGIADVIAGNIYWTQDQFVIEVARLNGSPRYV 2002  
QY 283 -SAGVDK- -T- - - - -I- -I- - 292  
Db 2003 ISQGLDKPRAITVHPEKGYLFWTEWQYPRIERSLDGTERVVLNVNVSISWENGISVDYQ 2062  
QY 293 - - - - -W-DA- - - - -H- - 296  
Db 2063 DGKLYCDARTDKIERIDLETGENREVLSSNNMDVSVFEDFIYWSDRTHANGSIKR 2122  
QY 297 - - - - -TG- - - - - 298  
Db 2123 GSKDNATDSVPLRTGIGVQDKIKVFNDRDROKGTNVCANGCQQLCLYRGGRACAC 2182  
QY 299 - - - - -E-A- - - - -K- -Q- - - - -F- 304  
Db 2183 AHGMLAEDGASCREYAGYLLYSERTILKSIHLSDERNLNAPVQPPDPPEHMKNVIALAFD 2242  
QY 305 - - - - -P- -F- -H- - - - -S- -A- - 309  
Db 2243 YRAGTSPGTPNRIFFSDIHFGNIQINDGSRITIVENVSGVEGLAYHRGWDTLWTSY 2302  
QY 310 - - - - -P-AL- - - - -D- -V-D- -W- -Q- - 317  
Db 2303 TTSTITRHTVDQTRGPAFERETVITMSGDHHPRAFLVLDQCNLMFWTNWNEQHPISIRAA 2362  
QY 318 -S- -N- - - - -T- - - - - 321  
Db 2363 LSGANVLTLEIKDIRTPNGLAIDHRAEKLYFSDATLDKIERCEYDGHSHRYVILKSEPVHP 2422  
QY 322 - - - - -F- - - - -A- - - - -SC- -S- - 326  
Db 2423 FGLAVYGEHIFWTDVVRRAVQRANKHVGNMKLLRVDIFQOPMGIIIVANDTNSCFLSPC 2482  
QY 327 - - - - -T- - - - -DM- - - - -CI- - - 331  
Db 2483 RINNGCQDLCLLTHGHVNCSCRGRIILQDDLTCEAVNSSCRAQDEFECEANGECINFLS 2542  
QY 332 - - - - -H- - - - -V- - - - -C-K 335  
Db 2543 TCDGVPHCKDKSDEKPSYCNRRCKTFRQCSNGRCVSNMLWCGADDGCGSDSEIPCNK 2602

QY 336 - - - - -LG- - - - -O- - - - - 338  
Db 2603 TAGGVGEPCRDGTCTIGNSSRCNQFVDCEDASDEMNCSATDCSSYFRLGVKGVLPQPCER 2662  
QY 339 - - - - -D- - - - -RP- - - - -I- - - - -K- - - - - 343  
Db 2663 TSLCYAPSVWCDGANDCGDYSDERDCPVKVRPRCLNYPACPSGRICPMSWTKCDKEDDCE 2722  
QY 344 - - - - -T- - - - -F- - - - -Q- - - - -G- - - - -H- - - - - 348  
Db 2723 HGEDETHCNKFCSEAQFECQNHRCISKQWLCGSDDCGSDGDEAAHCEGKTCGSPSFSCP 2782  
QY 349 - - - - -T- - - - -N- - - - -EV- - - - -N-A- -I- -K- - - - - 356  
Db 2783 GTHVCVPERWLCDGDKCADGADESAAGCLYNSTCDDREFMCONQCIPKHFVCDHDD 2842  
QY 357 - - - - -W- - - - -D- - - - -P- - - - -T- - - - - 360  
Db 2843 CADGSDSEPEYPTCGPSEFRCANGRCLSSRQWECDEGNDCHDQSDAPKNPHCTSPH 2902  
QY 361 - - - - -G- - - - -N- - - - -L- - - - - 363  
Db 2903 KCNASSQFLCSSGRVCAEALLCNGQDDCGSDSDERGHINECLSRKLSGCSQDCEDLKIG 2862  
QY 364 - - - - -L- - - - -A- - - - -S- - - - -CS- - - - -D- - - - - 369  
Db 2963 FKCRCPGFRKXDDGRTCADVDECSITTFPCSORCINTHGSYKCLCVGYAPRGDHPHCK 3022  
QY 370 - - - - -M- - - - -T- -LK- - - - -I- -W- - - - -SM 378  
Db 3023 AVTDEEPFLIFANRYVLRKLNLDGSGNYTLKQGLNNAVALDFYRQMIYVTDVTTQSGM 3082  
QY 379 - - - - -Q- - - - -D- - - - -N- - - - -C- - - - - 383  
Db 3083 IRRMHLNGSNVQLHRTGLSNPDGLAVDWVGNLYWCCKRDTIEVSKLNGAYRTVLVSS 3142  
QY 384 - - - - -V- - - - -H- - - - - 385  
Db 3143 GLREPRALVVDVQNGLYWTDGHSLSIGRMGSSRSVIVDTKITWPNGLTLDYVTER 3202  
QY 386 - - - - -D- - - - -L- -Q- - - - -AH- - - - - 390  
Db 3203 IYWADAREDIIEFASLDGSRHVLSDQIPHIFALTDFEDYVYVTDWETKSINRAHKTG 3262  
QY 391 -NK- - - - -EI- - - - - 395  
Db 3263 TNKTLILSTLHRPMDLHVHALRQPDVPHPCVNNNGGCSNLCILSPGGGHKACPTNFY 3322  
QY 396 - - - - -T- - - - -IK- -W- - - - -S- - - - -P- - - - - 401  
Db 3323 LGSdGRTCVSNCTASQFVCKNDKCIFFWVKCDTEDDCGHSDDEPPDCPFKCPGQFQCS 3382  
QY 402 TG- - - - -P- - - - -G- - - - -TN- - - - -N- - - - - 408  
Db 3383 TGICTNPAPICDGDNDQDNDEANCIDHVLCLPSQFKCTNTRNCIPGIFRCNGQDNCGDG 3442  
QY 409 - - - - -PN- - - - -ANL- -W- - - - - 414  
Db 3443 EDERDCPEVTCAPNQFQCSITKRCIPRVWVCDRNDVCDGSDDEPANCTQTCGVDEFRCK 3502  
QY 415 - - - - -L-A- - - - - 416  
Db 3503 DSGRCIPARWKCDGDDCGDSDGDEPEKCECDERTCEPYQFRCKNNRCVPGRWQCDYDNDG 3562  
QY 417 - - - - -S- - - - -A- - - - -F- - - - - 420  
Db 3563 DNSDEESCTPRPCSSSEFSCANGRCIAGRWKCDGHDGADGSDKDEKCTPRCMDQFQCKS 3622  
QY 421 - - - - -D-S- - - - -T-VR- - - - -L-W- - - - - 427  
Db 3623 GHCIPLWRKADADCMGSDGSEACGTGVRTCPLDFQCNTLCKPLAWKDGEDDCGDN 3682

QY 428 D-----V-----DRGIC-I-----HTLTK 440  
 Db 3683 SENEECARFCPPNRPFRKNDR-VCLWIGRQCDDGNDGDEEDCBPPTAHT-T- 3739  
 QY 441 H-----Q-----E-----P-----V 445  
 Db 3740 HCKDKKEFLCRNQRCLSSSLRCNMFDDCGDGSDEEDCSIDPKLTSCATNASIGDEARCV 3799  
 QY 446 -----Y-----S-----V-----A-F----- 450  
 Db 3800 RTEKAAYCACRSGFHTVPGQPCODINECLRFGTCSQLCNNTKGHLCSCAENFWKTHNT 3859  
 QY 451 -----S-----P-----D-----GR----- 455  
 Db 3860 CKAEGSEQVLYIADNNEIRSLFPGHPHSAVEQAQGDSESVRIDAMDVHVKAGVYWTNW 3919  
 QY 456 -----Y-L-A-S-----G-----S-----F----- 462  
 Db 3920 HTGTISYRSLPPAAPPTTSNRHRRQIDRGVTHLNISGLKMPRGTAIDWVAGNVYWTDSGR 3979  
 QY 463 D-----K-C-----V-H-I-----W-N-----T- 471  
 Db 3980 DVIEVAQMGENRKTLSGMDIPEHAIIVDPRLGRTMYWDMGNHPKIETAAMDGTLRETL 4039  
 QY 472 -----Q-TGAL-V-HS-Y-----R-GT-----G----- 484  
 Db 4040 VODNIQWPTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPDVAADSKRGLSHPPFSIDVF 4098  
 QY 485 -----G-----I-F-----EV-----C-W 491  
 Db 4099 EDYIYGVTYINNRVFKIHKFGHSPVLNLTGGLSHASDVLYHQHQPENVTPCDDKKCEW 4158  
 QY 492 -----N-----A-----A----- 494  
 Db 4159 LCLLSPGPVCTCNGKRLDNGTCVVPSPPTPPDAPRPGTCNLQCFNGSGCFINARRQP 4218  
 QY 495 -----GDK-----V-G-A- 500  
 Db 4219 KCRCQPRYTGDKCBLDQWEHCRNGGTCAASPSPGMPTRCPTGTPGPKCTQQVCAGYCAN 4278  
 QY 501 -----S-----ASDGS----- 506  
 Db 4279 NSTCTVQGNQPCRCPLGFLGRQYRCQSGYCNENFGTCQMAADGSRQCRCTAYFEGSR 4338  
 QY 507 -V-----CVL-----DL-----R-----K 514  
 Db 4339 CEVVKCSRLEGACVNVKQSGDYTCNCTDGRVAPSLCTCVGHCSNGSGCTMWSK 4392

## RESULT 3

US-09-334-220-2  
 ; Sequence 2, Application US/09334220  
 ; Patent No. 6323177  
 ; GENERAL INFORMATION:  
 ; APPLICANT: St. Jude's Children's Research Hospital  
 ; APPLICANT: Curran, Thomas  
 ; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
 ; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
 ; TITLE OF INVENTION: THERAPIES  
 ; FILE REFERENCE: 2427/05704  
 ; CURRENT APPLICATION NUMBER: US/09/334,220  
 ; CURRENT FILING DATE: 1999-06-16  
 ; NUMBER OF SEQ ID NOS: 5  
 ; SOFTWARE: FastSeq for Windows Version 3.0  
 ; SEQ ID NO 2  
 ; LENGTH: 3461  
 ; TYPE: PRT  
 ; ORGANISM: Mus musculus  
 US-09-334-220-2

Query Match 70.4%; Score 2583.8; DB 3; Length 3461;  
 Best Local Similarity 11.7%; Pred. No. 6.2e-33;

Matches 392; Conservative 93; Mismatches 16; Indels 2842; Gaps 359;  
 QY 1 M-----SI-----S-----D-----EV----- 8  
 Db 1 MERGCWAPRALVLAALLLTLARARATGYPPRPSFFFLCTHGELEGDEGEVULISL 60  
 QY 9 ---N-----F-LV-YRY---LQ---ES---GF---S-H 23  
 Db 61 HIAGNPYYVPGQBYHVHTISTSTFFDGLLVTGL-YTSTSIQSSQSIGSSAFGFGMSDH 119  
 QY 24 ---S-A-----F-----T-----F-----G-I----- 30  
 Db 120 QFNGQWCVSVASHVSHLPTNLFSFVWIAPPAGTCVNFMATATHRGQVIFDQALAQOLC 179  
 QY 31 ---E---SH-----I-----S-O---SNI-----N-----GAL----- 43  
 Db 180 EQCAPTEATAYSHLAETHSDVILRDDFDYSYQOLELNPNIWVECSNCEMGEQCQTTHMGN 239  
 QY 44 V---P-P-----A-AL---I-----SI-----IQ--- 54  
 Db 240 AVTFCEPYGPRELTTTCLNTTTASVLQFSIGSGCRFSYSDPSITVYAKUNTADMIQLE 299  
 QY 55 ---KG---LQY---V-EA-E-----VSIN----- 67  
 Db 300 KIRAPNSVTVIHLYPEEAKGESVQFQWKQDSLVRGEYACWALDNILV-INSARE 358  
 QY 68 ---ED---GT-LF-----DG-----R----- 76  
 Db 359 VVLEDNLDPDVTGNWLFPPGATVHSCQSDGNSIYFHGNEGSEFNFPATTRDVLSTEDIQ 418  
 QY 77 ---P---I-----ES-LSLI-----DA----- 86  
 Db 419 EOWSEFEFSQPTGWDILGAVWGADCGTVESGLSVFLKDGKRLCTPYMDTTGYGNLRFY 478  
 QY 87 -VM-----P-----D-----VV-----OT----- 94  
 Db 479 FVWGGICDPGVSHENDIILYAKIEGRKEHIALDTLYSSYKVPVSLVSVVINPELQTPATK 538  
 QY 95 ---RQ---QAY-R---D-----K-LAQ-----Q----- 106  
 Db 539 FCLRQKHOGYNNRVNAVDFFHVLVPLPSTMGHMIQFSINLGCCTHQPNGSVLSBFSTNH 598  
 QY 107 ---HA-----A-----AAA-AA-----AA-----AT--- 118  
 Db 599 GRWSLLHTCELPICAGPHLPHSTVSYSENYSYGNRNITIPLPNAALTRDTRIRWQTGP 658  
 QY 119 ---N-----Q-----Q-----Q-----GS 123  
 Db 659 ILGNMWAIDNVYIGPCLKFCSCGRQCTRHGCKCDPFGSPACEMASQTFPMFISESGS 718  
 QY 124 AK---N---G-E-----N----- 129  
 Db 719 ARLSYHNPFYSIRGAESVFGCGVLASGKALVFNKDRRLQTLITFLDSSQSREFLOFTLRG 778  
 QY 130 ---T-A-----NG---E-E---N-----G- 137  
 Db 779 SKSVLSTCRAPDQPGEGVLLHVSYDNGITWKLLEHYVYNVHEPRILISVELPDDARQGI 838  
 QY 138 ---A-H-----T-----I-----AN-----N-----H- 145  
 Db 839 QFRWQPYHSSQGEDVWAIDEIVMTSVLFNSISLDFTNLVEVTQSLGFLGNVQPCGHD 898  
 QY 146 ---T-D---M-----MEVDG-----D-V-E----- 156  
 Db 899 WTLCTGDSKLASSRMYVETOSMQI-GASYMIQFSLVMGCGQKYTPHMDNVKLEYSAH 957  
 QY 157 ---I---PSN---K-----A-----VVL-----R----- 166  
 Db 958 GLTWHLVQEECLPSPSCQEFTSASIVHASEFTQWRVTVVLPOKTWSGATFRWSQSY 1017  
 QY 167 ---GH-----E-SE-----V---F--- 173  
 Db 1018 TAQDEWALDNIIYIGQCCPNMCSGHGCDHGVCRCDQGYGQTECHPEALPSTIMSDFENP 1077

QY 174 -----I-----C-----A-----W----- 177  
Db 1078 SSWESDQWQVIGGVVKEGCGVSSGSSLYFSKAGKROLVSWDLDTSWDFVQFYQI 1137  
QY 178 -----N-P-----V-----SDL-----LV-----SG----- 187  
Db 1138 GGSSAACNKPDSREEGILLOYSNNGIOWHLLAEMVFSDFSKPRFVYVLELPAAGKTPCTR 1197  
QY 188 -----SG-----D-----S-----T-----A-----R-IR- 196  
Db 1198 FRWKKPVFSGEDYDQWAVDDIIILSKQKQVIVPVNPTLPQNYEKPAPFDPNQMNSVWL 1257  
QY 197 -----N-LSN-S-----T-----S-G-----P-----T-QL----- 209  
Db 1258 MLANEGMAKNDSCAFTPSAMVFGKSDGRFAVTRDLTLKPGYVLOFKNLIGTQSQFSS 1317  
QY 210 -----V-LR-H-----C-----IR-EG-----GQD-----V 222  
Db 1318 APVLLQYSHDAGMSWFLLEKGCFFASAAKCGEGRSRELSPTVYTG-DPEEWTRITIAI 1376  
QY 223 P-----S-N-----K-----D----- 227  
Db 1377 PRSLASKTRFRWIOESSOKNPPFGLDGVYISEPCPSYCSGHGDCISGVCFDLGYTA 1436  
QY 228 -----V-T-----S-LD-----W-----N-SEG-TL-----L-----A- 241  
Db 1437 AQGTCVSNTPNHSEMDFRPEGKSLPLWKITGGQVGTGCTLN-DGRSLYFNLGKREAR 1495  
QY 242 -----T-----GS-----Y-D-G-----F 248  
Db 1496 TVPLDTRNISLVQFYQIGSKTSGITVTPRAREYGLVQVSNNDGILWHLLRELFMSF 1555  
QY 249 -----A-RW-----T-----K-DG- 256  
Db 1556 LEQIISIDLPREAKTPATFR-WWQPHQKHAQWALGVILGVNDSSQTGFQDKLDGS 1614  
QY 257 -----N-----LAS-T-L-----G-Q-----H-----KG- 267  
Db 1615 IDLOANWYRLOQGOVDIDLCL-SMDTALIFTEIGNPRYAETWDFHVSSESLQWENMGC 1673  
QY 268 -----P-P-A-----L-----K-W-----NKK-----G----- 278  
Db 1674 SKP-FSGAHGILQYSLNKGQWLTVECPPTIGCVHYTESYTSYTSERFQWRRVTV 1732  
QY 279 -----N-----P-I-----I-----LSAGYDKT-TI-----I-----W-----DA----- 295  
Db 1733 LPLATNSPRTRFRWIOQNYTVGAD-SWADNVILASGCPMWCGRGICSDGRVCDRGFG 1791  
QY 296 -----H-----T-----GE----- 299  
Db 1792 GPFCVPVPLPSILKDDPFGNLHPDLWPEVYGAERGNLNGETIKSGTCLIFKGEGLRMLI 1851  
QY 300 -----AK-Q-----QF-----P-----F-----H-----S----- 308  
Db 1852 SRDLDCNTMYQVQSLRFAKGTPEKSHSILLQFSVSGVTVHLMDEFYPPQTTSLFIN 1911  
QY 309 AP-----A-----L-----D-----V-----D-W 316  
Db 1912 VPLPYGAQTNATRLWQPNYNGKBEIWIIDDFIIDGNLNNPVLILLDTDFGPRDNW 1971  
QY 317 -----S-N-N----- 321  
Db 1972 FFPYGGNIGLYCPSYKGAPEEDSAMVFSVNEVEHSITTRDLSVNEITIIQFEINVGCS 2031  
QY 322 -----F-A-----S-CST----- 327  
Db 2032 TDSSADPVRLEFRSDFGATWHLPLCYHSSSLVSLCSEHPSSTVYAGTQWRRE 2091  
QY 328 -----D-----MC-IH-----VC 334  
Db 2092 VWHFGKLHLCGVRFRWYQFYAGSPQVTVWADNVYIGQCCEMCMYGHGSCINGTKCIC 2151

QY 335 -----K-L-----GO-----DR-----P-----I-KT-----F-Q-G- 347  
Db 2152 DPOYSGPTCKISTKNPDLKDDPEGQLESDRFLMSGGKPSKCKGILSSGNLFFNEDGL 2211  
QY 348 -----H-----T-N-----EV----- 352  
Db 2212 RMLVTRDLDLSHARFVQFMRLGCGKGVPPRSPQVLLQYSLNGGLSWSLLQBLFNSNS 2271  
QY 353 NA-----I-----K-----W----- 357  
Db 2272 NVGRYTALEMPKARSSTRLRWQPSENGHFPVSPVVIDOILIGGNISGNTVLEDDPSTL 2331  
QY 358 D-----P-----TG-----N-----LL-----ASC-S-D 369  
Db 2332 DSRKLLHPGCTKMPVCGSTGDALVFEKASTRYVVTDIADVNEGSFLQIDFAASCSTVD 2391  
QY 370 -----DM-----T-----L-KI-----W----- 376  
Db 2392 SCVAIELEYSVDLGLSMHPLVRDCLPTNVECSRYHLQRLVSDTFNKTWTRITLPLPSYTR 2451  
QY 377 S-----M-----KQ-----DN-----CV-----H-----D- 386  
Db 2452 SQATRRFHPAPFDKQOTWADNVYIGDCLDMCSGRCVQSGCVDEQWGLYCDPE 2511  
QY 387 -----L-----Q-----A-H----- 390  
Db 2512 ETSPLTQLKDNFRAPSNQNLTVSGKLSLTVCGAVASGLALHFGSGCSRLLVTVLNLTL 2571  
QY 391 NKE-I-----Y-----T-----IKW-----S-PTG-----P 404  
Db 2572 NAEFIOFYMYGLITPFSNRNOGVLLSEYVNGGITWNLMEIFYDYQSKP-GFVNILLPP 2630  
QY 405 -----CT-----NPN-----N-A-NMLA-SA-----SF-----D 421  
Db 2631 DAKBIATRRWQ-PRHDGLDQNDWADNVILSGSADQRTVMDLTSSAPVPOHRSAPD 2689  
QY 422 S-TV-R-----L-----W-----DVD-----R-----GI-C-IH-----TL-TKH- 441  
Db 2690 AGPVGRIAFEMFLEDKTSVNEENWLPD-DCTVERFCDSPDGVMLCSGHDREVAVT-HD 2747  
QY 442 -----Q-----E-P-----V-YSVAF-----S-----P-D----- 453  
Db 2748 LPTENWIMQKISVCKVPEKIAQNIHVQFSTDFGVSWSYLVLPQCLPADPKCSGSVQ 2807  
QY 454 -----G-R-Y-L-AS-G-----S-----F-----D- 463  
Db 2808 PSVFPTEGWKRITYPLPESLTGNVFRFYQKYSVQWADNFYLPGLDNCGGHDC 2867  
QY 464 -K-CV-----H-----I-----W-----NT- 471  
Db 2868 LKEQCICDPGSGPNCYLTHSLKTLKRFDSBEIKPDLWMSLEGGSTCCEGVLAENTA 2927  
QY 472 -----Q-T-----GA-L-----V-----H-----S----- 479  
Db 2928 LYFGSGTVAITQDLDLRGAKFLQYWGRIGENNNMTSCHRPRVCRKEGVLLDFSTGGIT 2987  
QY 480 -----Y-R-----G-----T----- 483  
Db 2988 WTLHMDFOKYISVRHDYILLPEGALTWTRLRWQPPFVSNGLVSVGERAQWALDNI 3047  
QY 484 --GG-I-----P-----E-----V-----C-----W-----NA 493  
Db 3048 LIGGAEINPSQLVDTFDDSGSHEENWSFVPAVTRTAGFCGNPSPHLYWPKKKDKTHNA 3107  
QY 494 -----A-----GD-----K----- 497  
Db 3108 LSSRELIQGYMMQPKIVVGCEATSGDLHSMLEYTKDARSWSQLVQTCPLPSSNS 3167  
QY 498 VG-----A-----SA-----S-D- 504  
Db 3168 ICGSPQFHEATYNAVNSSSWKRTITIQLPDHVSSATQFRWIKQGEETEKQSWAIDHVV 3227  
QY 505 -G-----S-----VC----- 508



Db 1140 EKNCNSTETCQPSQFNHRCIDLFSVCDGDKDCVDGSDVGVCLNCTASQFKCASGDK 1199  
Qy 163 ---V---VL---R-G-H---E---169  
Db 1200 CIGVTRCDGVPCSDNSDEAGCPTRPPGMCHSDEFQCBQBDGICIFNFWECDDHPDCLYG 1259  
Qy 170 S-E-V---F---IC---AW---N---P---179  
Db 1260 SDEHNACVPKTCPSYFHCNGNCHIRAWLCDRDNDGDMDEKDCPTQFPFRCPSPWQWC 1319  
Qy 180 ---V---SD---182  
Db 1320 LGHNICVLSVWCDGIFDCPNGTDESPLCNGNSCDFNGGCTHECVQBPFGAKCLPLGF 1379  
Qy 183 LL---V---S---GS---G---189  
Db 1380 LLANDSKTCEDIDECIDLGSCSHCHYVMRGSFRCSCTGYMLESDBRTCKVTASESLLLL 1439  
Qy 190 ---DS-T---A---RI-W---N---197  
Db 1440 VASQNKIIADSVTSQVHNIYSLVNGSYIIVAVDFDSISGRIFWSDATQGTWSAFQNGTD 1499  
Qy 198 ---LSB---N---S---T---203  
Db 1500 RRVFPDSSIILTETIAIDWGRNLYWTDYALETIEVSKIDGSHRTVLISKNLTNPRGLAL 1559  
Qy 204 ---S-G---P-T---207  
Db 1560 DPRNEHLLFWSGHHPRIERASMDGSMRTVIVQKIFWPCGLTIDYPNRLLYFMDSYL 1619  
Qy 208 ---Q---LVLRH---213  
Db 1620 DYMDFCDYNGHRRQVIASDLIIRHPVALTLFEDSVYTTDRATRVMRANKWHGNGQSVV 1679  
Qy 214 ---CI-R---EG---G---219  
Db 1680 MYNIQWPLGIVAVHPSKQPNVNPFCAPSRCSHLCLLSSQGFHPFVSCVCPGWSLSPDLLN 1739  
Qy 220 ---Q---D-VP-S-N-K-D---227  
Db 1740 CLRDDQFLITVRQHIIFGISLNPVKSDNAMPVIAQIQLNGLDVEFDABEQIYIYVENPG 1799  
Qy 228 ---V---S---LDW---N---S-E---G---T 238  
Db 1800 EIHVRKTDGTRNVFASISVWGSMLALDWISERNLYSTNPTQSIIEVLTLHGDIRYKRT 1859  
Qy 239 LLA-T---G-S---Y---DG---F---248  
Db 1860 LIANDGTALGVGFPFIGITVDPARGLYWSQGTDSGVPAKIASANMDGTSVKTFTGNLE 1919  
Qy 249 ---A---R-I---W---T---253  
Db 1920 HLECVTLIDIEBQKLYWAVTGRGVIERGNVDGTRMILVHOLSHPWGIAVHDSFLYYTDEQ 1979  
Qy 254 ---K-D-G-N-L---ASTL-G-Q-H---K---266  
Db 1980 YEVIERVDRATGANKIVLRDNVNLRLGLQVYHRRNAESNGCSNNMACQIICLPVPGG 2039  
Qy 267 ---G---P---I-PALK---273  
Db 2040 LFSACATGKLPNDNRSCSPYNSFIWVMSLAIRGFSLELSDHSETMVPVAGQGRNALH 2099  
Qy 274 ---W---N-K---K---GN- 279  
Db 2100 VDVVSSGFIYWCDFSSVASDANAIIRIKPDGSSLNIVTHGIGENVRGIAVDWVAGNL 2159  
Qy 280 ---F---I---281  
Db 2160 YFTNAFVSETLIEVLINTYRRVLLKVTVMPRHIVVDPKRYLFWADYQRPKIERSF 2219  
Qy 282 ---L-S---A---G---VD---287  
Db 2220 LDCNTRTVLVSIGIVTPRGLAVDRSDGYVYVWDDSLDIIRINGENSEVIRYGRYPT 2279

Qy 288 ---K---T-T-I-I-W---D---294  
Db 2280 PYGITYFENSIIWDRNLKKIIOASKEPENTTEPTVIRDNINWLRDVTFDKQVQPRSPA 2339  
Qy 295 ---A---HT---G---E---A---300  
Db 2340 EVNNPCLENNGCGSHLCFALPGLHTPKDCAPGTLQSDGKNCAISTENELIFALSLSLR 2399  
Qy 301 ---K---Q---Q-F---P- 305  
Db 2400 SLHLDPENHSPPPTINVERTVNSLDYSDVSDRIYFTQNLASGVQISYATLSGIHTPT 2459  
Qy 306 ---F---306  
Db 2460 VIASGIGTAGTADFWITRIIYSDYLNQMINSMAGDSNRTVIARVPKPRAIVLDPDQCG 2519  
Qy 307 ---H---S-AP-ALD---313  
Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSLSVMPGLTLDYEEDLLYVWDASLQRIERST 2579  
Qy 314 ---VD---W---Q---S---N---319  
Db 2580 LTGVDRVIVNAVAHAFGLTYGYIYTDLYTQRIYRANKYDGSQGIAMTTLNLSQPRG 2639  
Qy 320 -NT---F---322  
Db 2640 INTVVKQKQCCNNPCEQFNGGCSHICAPGNCAECQCPEGNWLANNRKHICVIDNGER 2699  
Qy 323 -AS---CS---TD-M-C-IH---V---333  
Db 2700 CGASSFTCSNGRCISBEWKCDNDGDCGDESVICALHTCSPTAFTCANGRCVOYSYR 2759  
Qy 334 CKL---G---OD---R-P-I---342  
Db 2760 CDYNDGCGSDEAGCLFRDCAATFECNRRRCIPREFICNGVDNCHDNTSDEKNCPD 2819  
Qy 343 ---K---T---FQ---346  
Db 2820 RYCQSGYTKCHNSNICIPRYLDCGNDGCDNSDENPTYCTHTCSSEFQACASGRCIPO 2879  
Qy 347 ---GH---T---349  
Db 2880 HWYCDQETDCFDASDEPASCCHSERCTCLADEFKDCGRCIPSEWICDGDNDGDMSEDK 2939  
Qy 350 ---N---E---VNA---IK-W---357  
Db 2940 RHQCQONCSDEFLCVNDRPPDRRCIPOSWVCDGVDCTGDYDENQNCNTRTCSENEFT 2999  
Qy 358 ---357  
Db 3000 CGYGLCIPKIFRCDRHNDGCDYSDERGCLYQTCQOQFTCQNGRCISKTFVCDENDCGD 3059  
Qy 358 -D---PT---G---N-L---363  
Db 3060 GSDELMLHLCHTPEPTCPPEFKCDNGRCIEMMKCNHLDCLDCLNDSBKGGCINECHDPSI 3119  
Qy 364 ---LAS-CS---D-D-M---371  
Db 3120 SGCDHNCTDLTSLFYSCSPGYKLSMDKRTCDVIDECTEMPFVCSQKCNENVISYICKCA 3179  
Qy 372 ---T---L---374  
Db 3180 PGYLREPKDTCQNSNIEPYLIFSNRYLRLNLTIDGYFSLILEGLDNVVALDFOREK 3239  
Qy 375 ---I---W---S---377  
Db 3240 RLYWIDTQRIERMFNKTNETIINHRLPAESLAVDWVSRKLYWLDARLDGLFVSDL 3299  
Qy 378 ---M-KQ---D-N-C---VHDLQ---A-H---NK---392  
Db 3300 NGHREMLQACHCVDANNTFCFONPRGLALHP-QYGYLYWADWGHRAIYIGRVGMDGNTKSV 3358

QY 393 -----E-----I-YT-----IK----- 398  
Db 3359 IISYKLEWPNGITDYTDNLLYWAHLYIEYSDLEGHHRHTYVDGALPHPPFAITIFED 3418  
QY 399 -----W-----S-P-T-G--- 403  
Db 3419 TIYTDNTRTEKGNKYDGNRQTLVNTTRPFDIHVYHPYRQPIVSNPGTNGGCSH 3478  
QY 404 -----PG-----T-----NN----- 408  
Db 3479 LCLIKPGKGFTCEPDDFRTLQLSGSTYCMPMCSSTQFLCANNEKCIPIWKKDQKDC 3538  
QY 409 -----P-----NA-N-----LML-----AS- 417  
Db 3539 SDGDELALCPQRCRLGQFCSDGNCTSPOTLCNAHQNCPDGSDERL-LCENHHCDSN 3597  
QY 418 -----A-----S-----F-----DS-----T-----VR-----L-----W-DV- 429  
Db 3598 EWQCANRKCIPESWQCOTFNDCEDESDSHCASRTRCPQGFRCANGRCIPQAWKCDVD 3657  
QY 430 -DRG-----I-CI-----H-----T-L----- 438  
Db 3658 NDCGDSDEPTEECSSAHLCDNFTFSCNTYNYRCPKWA/CNGVDDCRDNDSDQGCER 3717  
QY 439 T-----K-H-----Q-----E-P-----V----- 445  
Db 3718 TCHPVGDFRCKNHHCIPLRWCDCQNDGNDSDENECAPRECTESEFRVCVNOQCIPSRWI 3777  
QY 446 -----Y-----S-V-----A-----FS 451  
Db 3778 CDHYNDGDSNDRDCERMTCHPEYFQCTSHCVHSELKCDGSADCLDASDEADCPTRF- 3836  
QY 452 PDGRL-AS-----GS-----F----- 462  
Db 3837 PDGAYCATWFECKNHVCIPPYWKCDGDDCGDGBELHLCLDVPNCNRRFRCDNNRC 3896  
QY 463 -----D-----K-----CV-H--I-----W 469  
Db 3897 IYSHEVCNGVDDCGDGTDETEHCKRPTPKPCTEYKCGNGHCIPHDNVCDADDGCDW 3956  
QY 470 -----N-----T-----Q-T-----G-----A-----L-V----- 477  
Db 3957 SDELGCNKGKERTCAENICQNCQTLNEGGFICSTAGFETNVFDRTSCLDINECEQFGT 4016  
QY 478 -----H-----SY-----W-----N-----AA----- 481  
Db 4017 CPQHCRTNKGSYECVACADGFTSMRDKRCAEAGSGSPLLLLPDNRVIRKYNLSSEFSE 4076  
QY 482 -----G-----T-----G-----I--FE-----V-C--- 490  
Db 4077 YLQDEEYIQAVDYDWDPKDGLSVVYTVRGEGRFGAIKRAVIPNFESGRNNLVQEVDL 4136  
QY 491 -----W-----N----- 494  
Db 4137 KLYVMQPDGIADVWGRHYIWSVKNRIEVAKLDRYRKNLSTLDQPPAAIAVNPXL 4196  
QY 495 -----D-----K-V----- 498  
Db 4197 GLMFWDWGKEPKIESAMNGEDRNILVPFDLGNPTGLSIDYLNDRYWSDFKEDVIETI 4256  
QY 499 -----G-----A-SA-----S-----D-----G----- 505  
Db 4257 KYDGTDRRVIAKEAMNPYSIDIFEDQLYIWSKEGVEWKNQKFGQKKEKTLVNPWLQ 4316  
QY 506 -----SV-----C-V-----L-----D-----L 512  
Db 4317 VRIFHLRYKSVFNLKQICSHLCLLRPGYSCACPQSSFTGSGTTECDAAIELPINL 4376  
QY 513 -----R-----K 514  
Db 4377 PPRCRMHGNGCYFDETDLPK 4397

## RESULT 5

US-08-652-877-84  
; Sequence 84, Application US/08652877  
; Patent No. 6187548  
; GENERAL INFORMATION:  
; APPLICANT: Akerstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Rask, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjaln, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
; TITLE OF INVENTION: Thereof and DNA Encoding Same  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd., 3C43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.5.1  
; SOFTWARE: Word 6.0 (Patentin)  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/652,877  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA: PCT/US95/15203  
; APPLICATION NUMBER: 22-NOV-1995  
; FILING DATE: 23-NOV-1994  
; APPLICATION NUMBER: US 08/344,836  
; FILING DATE: 07-JUNE-1995  
; APPLICATION NUMBER: US 08/487,314  
; FILING DATE: 29,699  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Savitzky, Martin  
; REGISTRATION NUMBER: A1355E-US  
; REFERENCE/DOCKET NUMBER:  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 610-454-3816  
; TELEFAX: 610-454-3808  
; INFORMATION FOR SEQ ID NO: 84:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4655 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-652-877-84

Query Match 70.3%; Score 2577.8; DB 3; Length 4655;  
Best Local Similarity 9.8%; Pred. No. 4e-32; Indels 3892; Gaps 381;  
Matches 431; Conservative 59; Mismatches 20;

QY 1 M-----S-----I-----S--SDE- 7  
Db 1 MDRGPAAVACTLLALVACLAPASGQECDSAFHRCGSGHCI PADWRCDGTGKDCSDDAEI 60  
QY 8 ---VN---F---L---V----- 12  
Db 61 GCAVTVCOQGYPKQSEGGCIPSSWVCDQDQDDGSDERQDCSQSTCSSHQTCSNGQC 120  
QY 13 ---YR-----Y-----L-----Q-----E---S 19  
Db 121 IFSEYRCDHVRDPCPDGADENDCOYTCBOLTCDNGACYNYSKCDWKVDCRSDSDEINCT 180  
QY 20 -----G-FS-----H-----S---A--F-T-----G--I----- 30



Db 181 EICLHNEFCGNGECIPRAYVVDHNDQDGSDEHACNYPTCGGYQFTCPBGRCIYQNWV 240  
QY 31 -----ES-----H-----IS-----Q-SNI 39  
Db 241 CDGEDCKONGDEGGESGPHVHKSPREWSPEGRGCIYKVCVGDGILDCGRDEN - 299  
QY 40 N---G---AL---VP---PAALJ---S-----I-I-- 53  
Db 300 NTSTGKYCSMTLCSALNCQYQCHETPVGGACFPGPVLIINHDSRTCVEFDDCQINGICD 359  
QY 54 QK-----G-L--QY-----VE 61  
Db 360 QKCESRGRHLCHCEBYILRGYCKANDSPGEASIIFNSGRDLLIGDIHGRSPRIIVE 419  
QY 62 -----A-----E-VSI-----N 67  
Db 420 SQNRGVAVGAFHYHLQRFVWTQVQKVFSDVINGLNIQEVLVNSVETPENLAVDWVN 479  
QY 68 -----E-----DG---TL----- 72  
Db 480 KIYLVETKVNRIWMVNLGDSYRVTLITENLGHPRGIAVDPTVGYLFFSDWESLSGPKLE 539  
QY 73 --P-DG-R-----P-----IE-----S 80  
Db 540 RAFMDGSRKDLVKTKLGNWPAGVTLDMISKRVYWDVSRFDYIETVYDGIQRTVVHGG 599  
QY 81 L-----SL-----I-D-AVM----- 88  
Db 600 LIHPHFGVSLFEGQVFTDTKMAVLKANKFTETNPQVYQASLRPYGVTVYHSLRQPYA 659  
QY 89 --P-D-----V-V-----Q-T----- 94  
Db 660 TNPCKDNNGGCEQCVLSHRTDNDGLGFRCKCTFGQLDTERHCAVQNFILFSSQVAI 719  
QY 95 R-----Q-----Q----- 97  
Db 720 RGIPFTLSTQEDVMVPSGNPSPFVGIDFDAQDSTIFFSDMSKHMIFKQKIDGTGREILA 779  
QY 98 -----A-----Y--R-DK-LAQ-----Q--H-A-- 108  
Db 780 ANRVNVESLAFDWISKNLWTDSHYKSIWMLADKTRTVOYLYNNPRSVVHPFAGY 839  
QY 109 -----A-----A-----AA-----A----- 114  
Db 840 LPFTDFRPKIMRAWSDGSHLLPVNTTLGWENGLAIDWAASRLYWDAYPDKIEHSTP 899  
QY 115 -----A-----A-----A-----T----- 118  
Db 900 DGLDRRLGHIEQWTHPFGLAIFGEHLFTTDWRLGAIIRVRKADGGEMTVIRSGIAYILH 959  
QY 119 -----N-OQGS-A-----KNG-----EN-----TA-N-----G----- 133  
Db 960 LKSYDNIQTSNACNPTHPNGDCSHFCPPVFNFORVCGPYGMRLASNHLTCBGDPTN 1019  
QY 134 E---E-----N-----GA-H-----T----- 140  
Db 1020 EPTEQGLFSPFKNGKRCVPNYLDCGVDDCHDNSDEQLCGTLNNTCSSAFTCGHGEC 1079  
QY 141 I-A-----N-----H-----T-DMME-----V-D-----GD----- 154  
Db 1080 IPAHWRCKDENKDCVDSDEHNCPTHAPASCLDTQYTCDNHQICISKWVCDTNDNDCGDS 1139  
QY 155 -----VEI--PS-----N-----K-A----- 162  
Db 1140 EKNCNSTETQPSQFQNCNPHRCIDLSPVCDGDKDCVDGSDVGVCLNCTASQPKCASGDK 1199  
QY 163 -----V-----VL-----R-G-H-----E----- 169  
Db 1200 CIGVTNRCDGVFDCSDNSDEAGCTPRPGMCHSDEFCQEDGICIPNFWECDHGPDCLYG 1259  
QY 170 S-E-----V-----F-----IC-----AW-----N-----P----- 179  
Db 1260 SDEHNACVPKTCPSYFHCNNGNCIHRAWLCDRDNDGCMDSDEKDCPTQPFRCPSQWOC 1319

QY 180 -----V-----SD----- 182  
Db 1320 LGHNICVNLVVCDDGIFDCPENGTDSPCLNGNSCDFNGGCTHECVQEPGAKCLCPG 1379  
QY 183 LL-----V-----S-----GS-----G----- 189  
Db 1380 LLANDSKTCEDIDECIDILGSCSOHCYNMRGSRFCSDCTGYMLESDDGRTCKVTASESILL 1439  
QY 190 -----DS-T-----A-----RI-W-----N----- 197  
Db 1440 VASQNKIADSVTSQVHNIYSLVNGSYIVAVDFDISGRIFWSDATQGTWAFQMGTD 1499  
QY 198 -----LSE-----N-----S-T----- 203  
Db 1500 RRVVPDSSIIITETIAIDVGNRLYWDYALETIEVSKIDGSHRTVLISKNLINPRGLAL 1559  
QY 204 -----S-G-----P-T----- 207  
Db 1560 DPRMNEHLLFWSDGWHHPRIERASMDGSMRTVIQDKIFWPCCGLTIDYPNRLLYFMDSYL 1619  
QY 208 -----Q-----LVLRH----- 213  
Db 1620 DYMFCDYNGHRRQVIASDLIIRHPYALTLPEDSVYWTDRTRVRMRANKWHGGNQSVV 1679  
QY 214 -----CI-R-----EG-----G----- 219  
Db 1680 MYNIQWPLGIVAVHPKQNSVNPFCASRCSHLCLLSSQGPHEFVSCVPSGWSLSPLDLN 1739  
QY 220 -----Q-----D--VP-S--N-----K-D----- 227  
Db 1740 CLRDDQPLITVRQHIIFGISLNPVKSDAMVPIAGIQNGLDVEFDABQYIYVVENPG 1799  
QY 228 -----V-----T-----S-LDW-----N--S-E--G-----T 238  
Db 1800 EHRVKTGTNRVTFASISMVGPSMNALDWSIRNLSTNPTQTSLEVLTLHGDIYRKT 1859  
QY 239 LLA---T-----G-S-----Y-----DG-----F----- 248  
Db 1860 LIANDGTALGVGPPIGITVDPPARKLYWSDQGTDSVPKASIANMDGTSTVKLTG 1919  
QY 249 -----A-----R-I-----W-----T----- 253  
Db 1920 HLECVTDIEQKLYWAVTGRGVIERGNVDGTDMLVHQLSHPWGIAVHDSFLYTTDEQ 1979  
QY 254 ---K-D--G-N--L---ASTL-G-Q--H-----K----- 266  
Db 1980 YEVIERVKATGANKIVLRDNPVNLRLGLQYHRRNAEBSNGCSNNMNACQICLPVPGG 2039  
QY 267 -----G-----P-----I--FALK----- 273  
Db 2040 LFSACATGFKLNPDRNSCPYNSPIVVSMLSAIRGFSLELSDHSETWVPVAGQGNALH 2099  
QY 274 -----W-----N-----K-----K-----GN- 279  
Db 2100 VDVVSSGFIYWCDFSSVASDNAIRRIKPDGSSLMNIVTHGICENGVRGIAVDWVAGN 2159  
QY 280 -----P-----F-----I----- 281  
Db 2160 YFTNAFVSTLIEVLRINTYRVLKVTVDMPRHIVDPKRYLPWADYGQRPKIERSP 2219  
QY 282 -----L-S-----A-----G---VD----- 287  
Db 2220 LDCNRTVLVSEGITVTPRGLAVDRSDGYVYVVDVDSLDIARIRNGENSEVIRGSRPT 2279  
QY 288 -----K-----T-----T-I--I-W-----D----- 294  
Db 2280 PYGITVFENSIWVDRNLKKIIFOASKEPENTPTVIRDININWLRDVTIFDKQVQPRSPA 2339  
QY 295 -----A-----HT-----G-----E-----A----- 300  
Db 2340 EVNNPCLNNGGCSHLCPALPGLHTPKCDCAFGLTQSDGKNCAISTENFLIFALSNSLR 2399

301 QY -----K-----Q-----Q-F-----P- 305  
2400 Db SLHLDPENHSPFOTINVERTVMSLDYSDSDRIYFTONLASGVQISYATLSSGHTPT 2459  
306 QY -----F----- 306  
2460 Db VIASGIGTADGIAFDWITRRIYSDYLNQMSMAEDGSNRTVIARVPKRAIVLDFCQG 2519  
307 QY -----H-----S-AP-ALD----- 313  
2520 Db YLYWADWDTHAKIERATLGNFRVPIVNSLVMPGSLTLYEEDLLYWDASLQRIERST 2579  
314 QY -----VD-----W-----Q-----S-----N----- 319  
2580 Db LTGVREVIINAAVHAFGLTYGYIYTDLTYORIYRANKYDGSQGIAMTTLNLSOPRG 2639  
320 QY -----NT-----F----- 322  
2640 Db INTVVKNQKQCNPCQFNGGCSHICAPGNGAECOCPEHGNWYLANNRKHCIVDNGER 2699  
323 QY -----AS-----CS-----TD-M-----C-IH-----V----- 333  
2700 Db CGASSFTCSNGRCISEBWKCDNDNDGSDGSDMESVICALHTCSPTAFTCANGRCVQYSYR 2759  
334 QY -----CKL-----G-QD-----R-P-I----- 342  
2760 Db CDYNDGCGSDGAGCLFRONATTEPMCNRRICIPREFICNGVDNCHDNTSDKNCPS 2819  
343 QY -----K-----K-----T-----FQ----- 346  
2820 Db RTCQSGYTKCHNSNICIPRVVLCDGNDGSDNENPTYCTHTTCSSEFQCSAGRCIPQ 2879  
347 QY -----GH-----T----- 349  
2880 Db HWYCDQETDCEPASGCSHRTCLADEFKDGRGRCIPSEWICDGDNDGSDMSDEK 2939  
350 QY -----N-----E-----VNA-----IK-----W----- 357  
2940 Db RHQONQNCSDSFLCVNDRPPRCIPQSWVCDGVDCTDGYDENQNCNTRTCSNEFT 2999  
358 QY ----- 357  
3000 Db CGYGLCIPKIFRCDRHNDGSDYSDERGCLYQTCQOQFTCQNGRCISKTEVCDEDNDGSD 3059  
358 QY -----D-----PT-----G-----N-----L----- 363  
3060 Db GSELMHLCHTPTCPPEHFKCDNGRCIEBWKLCNHLDCDLSNDEKGCINECHDPSI 3119  
364 QY -----LAS-----CS-----D-D-----M----- 371  
3120 Db SGCDHNCITLTSFYCSRCRGYKMSDKRTCVDIDECETEMPFVCSQKCNVIGSYICKCA 3179  
372 QY -----T-----L----- 374  
3180 Db PGVLRPDGKTQRQNSNIEPYLIFSNRYRLNLTIDGYFYSILEGLDNVVALDFRVEK 3239  
375 QY -----I----- 377  
3240 Db RLYWIDTQROVIERMFLANKTKETIINHLRPAESLAVDWVSRKLYWLDARLDGLFVSD 3299  
378 QY -----M-KQ-----D-N-----C-----VHDLQ-----A-----H-----NK----- 392  
3300 Db NGGHRMLAQCHVDANNFTCFDNPGRGLAP-QGYLYWADWGHRAVIGRVGMDGNTKSV 3358  
393 QY -----E-----I-YT-----IK----- 398  
3359 Db IISTKLEWPNGITIDYTNLLYWADAHGLYIEYSDLEGGHRTVYDGLPHFPFAITIFED 3418  
399 QY -----W----- 403  
3419 Db TIYTWNTVTEKGNKYDGSNRQTLVNTTHRPFDIHVHPYQPIVSNPCGNNCGCSH 3478  
404 QY -----PG-----T-----NN----- 408

3479 Db LCLIKPGGKFTCECPDDFRTLQLSGSTYCMPCMSSTQFLCANNEKCIPIWKKDQKDC- 3538  
409 QY -----P-----NA-----N-----LML-----AS- 417  
3539 Db SDGSDALALCPQRFCLRGQFCSDGNCSTPQTLNAHQNCPCDGSDEDL- LCNHHCDN 3597  
418 QY -----A-----S-----F-----DS-----T-----VR-----L-----W-----DV- 429  
3598 Db EQCANKRCIPESWQCDTDFNDCEDNDESDSHCASRTCPGQFCRANGRCIPQAWKCDVD 3657  
430 QY -----DRG-----I-----CI-----H-----T-----L----- 438  
3658 Db NDCGDSHDEPIECMSAHLCDNFTEFSCKTNTYRCPKAVCNVGVDDCRNSDEQCEER 3717  
439 QY -----T-----K-----H-----Q-----E-----P-----V----- 445  
3718 Db TCHPVGDFRCKNHHCIPLRWQCDGQNDGSDNDEENCAPECTESBEFCVNOQCIPSRWI 3777  
446 QY -----Y-----S-----V-----A-----PS- 451  
3778 Db CDHYNDGSDNDRDCMRTCHPEYFQCTSGHCVHSELKDGSDADCLDASDEADCTRF- 3836  
452 QY -----PDGRL-----AS-----GS-----F----- 462  
3837 Db PDGAYCOATWFECKNHVICIPPYWKCDGDDDDGSDDEELHLCLDVPNSNPRFRCDNRC 3896  
463 QY -----D-----D-----K-----CV-H-----I-----W- 469  
3897 Db IYSHEVCNGVDDCGDGTDETEHCRKPTPKPCTEYKCGNGHICI PHDNVCDADDGCDW 3956  
470 QY -----N-----T-----Q-----T-----G-----A-----L-V----- 477  
3957 Db SDELGNKGKERTCAENICEQNCQTLNEGFGICSTAGFETNVFDRTSCLDINECEQFGT 4016  
478 QY -----H-----SY----- 481  
4017 Db CPOHCNTKSGYECVCADGFTMSDRPGKRCAGSSPILLPLPDNVIRKYNLSSEFSE 4076  
482 QY -----G-----T-----G-----G-----I-----FE-----V-C----- 490  
4077 Db YLQDEBYIOAVDYDWDPKDIGLSVVYTVRGEGRFGAIKRAYIPNFESGRNNLVQEVDL 4136  
491 QY -----W-----N-----AA----- 494  
4137 Db KLKYNQPGIAVDWVGRHIYNSDVKNKRIEVAKLGRYRKWLITDLDQPAIAVNPKL 4196  
495 QY -----G-----G-----D-K-----V----- 498  
4197 Db GLMFWTDWKEPKIESAWNGEDRNILVFDLGNWPTGLSIDYLNNDRIYWSDFKEDVIET 4256  
499 QY -----G-----A-SA-----S-----D-----G----- 505  
4257 Db IKYDGTDRRVIKAEAMNPYSLDIFEDQLYWSKEGEVKNQKFGQKKEKTLVNPWLT 4316  
506 QY -----SV-----C-V-----L-----D----- 511  
4317 Db QVRIFHQLRYNKSVPNLCKOICSHLCLRLPGYSGACPOGSSFIEGSTTECDAAILPIN 4376  
512 QY -----L-----R-----K- 514  
4377 Db LPPPCRCMHGNGCYFDETDLPK 4398

## RESULT 6

US-08-652-877-88  
; Sequence 88, Application US/08652877  
; Patent No. 6187548  
; GENERAL INFORMATION:  
; APPLICANT: Akerstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Rask, Lars  
; APPLICANT: Crumley, Gregg R.



QY 190 ---DS-T---A---RI-W---N---197  
Db 1440 VASOKIIADSVISOVHNIYSLVNGSYIIVAVDFDSISGRIFWSDATOGKTWAFQNGTD 1499  
QY 198 ---LSE---N---S-T---203  
Db 1500 RRVVFDSSIIILTETIAIDWGRNLWTDYALETIEVSKIDGSHRTVLISKNLTPRGAL 1559  
QY 204 ---S-G---P-T---207  
Db 1560 DPRMNEHLLFWSDGHHPRIERASMDGSMRTIVQDKIFWPCGLTIDYPNRLLYFMDSYL 1619  
QY 208 ---Q---LVLRH---213  
Db 1620 DYMDPCDYNHRRQVQIASDLIIRHPYALTLPEDSVYVWTRATRVVRANKWGHGNSVV 1679  
QY 214 ---CI-R---EG---G---219  
Db 1680 MYNIQWPLGIVAVHPKQPNVNPAPSRCSHLCLSSQGFHFVSCVCPGWSLSPDLLN 1739  
QY 220 ---Q---D---VP-S---N---K-D---227  
Db 1740 CLRDDQFLITVQHIIFISLNPVKSNDAMPVIAIGIQLDVEFDABEQYIYVVENPG 1799  
QY 228 ---V---T---S---LDW---N---S-E---G---T 238  
Db 1800 EIHVRKTDGTRNVFASISVWGPMSMLALDWISRNLYSTNPTQSEVLTLHGDIYRKT 1859  
QY 239 LLA---T---G-S---Y---DG---F---248  
Db 1860 LIANDGTALGVGPIGITVDPARGKLYWSQDTSQGVPAKIASANMDGTSVKLTFTGNLE 1919  
QY 249 ---A---R-I---W---T---253  
Db 1920 HLECVTLDIEQKLYMAVTRGVIERGNVDGTRMILVHQLSHPWGIAVHDSFLYTDEQ 1979  
QY 254 ---K-D---G-N---L---ASTL-G-Q-H---K---266  
Db 1980 YEVIERVXKATGANKIVLRDNVFNRLGLQVYHRRNAESSNGCSNNACQOICLPVPGG 2039  
QY 267 ---G---P---I---FALK---273  
Db 2040 LFSCACATGKLPNDNRSCSPYNSFIWVMSLAIRGFSLESDHSETWVPVAGQGRNALH 2099  
QY 274 ---W---N---K---K---GN- 279  
Db 2100 VDVVSSGFIYWCDFSSVASDRAIRRIKPDGSSLMNIVTHGENGVRGIAVDWVAGNL 2159  
QY 280 ---F---I---281  
Db 2160 YFTNAFVSETLIEVLRINTYRRVLLKVTVDMPRHIVVDPKNRYLFWADYQRPKIERSF 2219  
QY 282 ---L-S---A---G---VD---287  
Db 2220 LDCTNRTVLVSEGIIVTPRGVLAWSRSGVYVWVDDSLDIARIINGENSEVIRYSGRYPT 2279  
QY 288 ---K---T---T---I---I-W---D---294  
Db 2280 PYGITVFENSIIWDRNLKIFQASKEPENTPPTVIRDINWLRDVTIFDKQVQPRSPA 2339  
QY 295 ---A---HT---G---E---A---300  
Db 2340 EVANNPCLENNGGCSHLCPALPGLHPKCDCAFGTLQSDGKNCAISTENFLIFALNSLR 2399  
QY 301 ---K---Q---F---P---305  
Db 2400 SLHLDPENHPPPTQINVERTVMSLDYSDVSDRIYFTQNLASGVQGISVATLSSGIHTPT 2459  
QY 306 ---F---306  
Db 2460 VIASIGTAGDIAPWITRIIYSDYLNQMINMAEDGSNRTVIARVPKRAIVLDPCCQ 2519  
QY 307 ---H---S---AP---ALD---313

Db 2520 YLYWADWDTHAKIERATLGNFRVPIVNSSLVMPSGTLTLDYEEDLLYVWDASLQRIERST 2579  
QY 314 ---VD---W---Q---S---N---319  
Db 2580 LTGVDRREVIVNAVAHAGLTLYQYIYVTDLYTORIYRANKYDGSQGIAMTTNLLSQPRG 2639  
QY 320 ---NT---F---322  
Db 2640 INTVKNQKQCCNNPCEQFNGGCSHICAPGPNBAECQCPEGHNYLANNRKHICVNGER 2699  
QY 323 ---AS---CS---TD-M---C-IH---V---333  
Db 2700 CGASSFTCSNGRCISEWKCDNDNDGDSDESVICALHTCSTPTAFTCANGRCVQYSYR 2759  
QY 334 CKL---G---QD---R-P-I---342  
Db 2760 CDYNDGCGSDRAGCLFRDCNATTEFCMNRRCIPREFICNGVDNCHDNNTSDEKNCPD 2819  
QY 343 ---K---T---T---PQ---346  
Db 2820 RTCQSGYTKCHNSNICIPRYLDCGNDGCDNSDENPTYCTHTCSSEFQCASGRCIPO 2879  
QY 347 ---GH---T---349  
Db 2880 HWYCDQETDCFADSPASCGHSERTCLADEFKDCGRCIPSEWICDNDGCDMSDEK 2939  
QY 350 ---N---E---VNA---IK-W---357  
Db 2940 RHOCQONQCSDFLFCVNDPRPDRRCIPOSQVWCGDVDCDGDYDNGQNCNCTRTTCSNEFT 2999  
QY 358 ---358---357  
Db 3000 CGYGLCIPKIFRCDRHNDGDSYDERGCLYQTCQOQFTCQNGRCISKTFVCDENDCGD 3059  
QY 358 ---D---PT---G---N---L---363  
Db 3060 GSDLMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCNHLDDCLNDSDEKKGGINECHDPSI 3119  
QY 364 ---LAS---CS---D-D-M---371  
Db 3120 SGCDHNCTDLTSLFYSCRPYKLSMDKRTCDVIDECTEMPFVCSQKCNENGVISYICKCA 3179  
QY 372 ---T---L---K---374  
Db 3180 PGYLREPDGKTCRQNSNIEPYLIFSRYLRLNLTIDCYFYSLLILEGLDNVVALDFRVEK 3239  
QY 375 ---I---W---S---377  
Db 3240 RLYWIDTQVIERMFLNKTNETIINHRLPAESLAVDWVSRKLYWLDARLDGLFVSDL 3299  
QY 378 ---M-KQ---D-N-C---VHDLO---A-H---NK---392  
Db 3300 NGHRRMLAQHCVDANNTFCFDPNPRGLALHP-QYGYLYWADWGHRAIYGRVGMGTNKS 3358  
QY 393 ---E---I-VT---IK---398  
Db 3359 IISTKLEWNGITIDYNDLLYWADAHLYEYSDLEGGHRRHTVYDGLALPHPAITIFED 3418  
QY 399 ---W---S-P-T---G---403  
Db 3419 TIYWTDMNTRTVEKGNKYDGSNRQTLVNTTHRPFDIHVYHPYRQPIVSNPCGTNNGCSH 3478  
QY 404 ---PG---T---NN---408  
Db 3479 LCLIKFEGKGFTCECPDPFRTLQLSGSTYCMPCMSSTQFLCANNEKCIPIWKKCDGQKOC 3538  
QY 409 ---P---NA-N---LML---AS- 417  
Db 3539 SUGSDEALCPORFCKLQFQCSGDNCTSPQTLCAHQNCPOGSDERL-LCENHHCDSN 3597  
QY 418 ---A---S---F---DS---T---VR---L---W-DV- 429

Db 3598 EWQANKRCIPESWQCTFNDCEDESDSHCASRTCRPGQFRCANGRCIPOAWKCDVD 3657  
Qy 430 -DRG-----I--CI--H-----T-L-----438  
Db 3658 NDCGDHSDPIECMSAHLCDNFTEBSCNTNRPKWAVCNVDDCRDSDQGEER 3717  
Qy 439 T-----K--H-----Q-----E--P-----V-----445  
Db 3718 TCHPVGDFRCKNHHCIPLRWQCDGNDGDSBDEBENAPRECTESBPRCVNQOCIPSRWI 3777  
Qy 446 -----Y--S--V-----A-----FS 451  
Db 3778 CDHYNDCGDSRDERDCMRTHCEYFOCTSGHCVHSELKCDGSADCLDASDEADCPTRF- 3836  
Qy 452 PDGRYL-AS-----GS-----F-----462  
Db 3837 PDGAYCOATPECKNHVCIPPYKWKCGDDGDSDEELHCLDVPNSPNNRCNNRC 3896  
Qy 463 -----D-----K-----CV-H--I-----W 469  
Db 3897 IYSHVCGVDDCGDGTDETEEHCKPTEYKCGNGHCIPHDNVCDADDGCDW 3956  
Qy 470 -----N-----T-----Q--T-----G-----A-----L-V-----477  
Db 3957 SDELGNKKGKERTCAENICQNCQTLNNEGFCISCTAGFTNVFDRTSCLDINECQFGT 4016  
Qy 478 ---H-----SY-----T-----G-----I--FE-----V-C---490  
Db 4017 CPQHCRTKGSYECVADGFTMSDRPKCAEGSPLLLDPNVIRKYNLSERFSE 4076  
Qy 482 -----G-----T-----G-----I--FE-----V-C---490  
Db 4077 YLQDEEYIAVDYDWDPEDIGLSVWYTVRGEGRFGAIAKRAVIPNFESGRNNLVQEVDL 4136  
Qy 491 -----W-----N-----AA-----494  
Db 4137 KLKYNQPDGIADVWGRHIYSDVKNKRIEVAKLGRYKWLISTDLQPAIAVNPXL 4196  
Qy 495 -----G-----G-----D-K--V---498  
Db 4197 GLMFWTDWGEKPKESANWNGEDBNILVFDLQWPTGLSIDYLNDRYISDFKEDVIET 4256  
Qy 499 ---G-----A-SA---S---D-----G-----505  
Db 4257 IKYDGTDRRVIKAEAMYPYSLDIFEDQLYWISKEGEWKNKFGQKKEKTLVVPWLT 4316  
Qy 506 -----SV---C-V---L-----D-----511  
Db 4317 QVRIFHQLRYNKSVPNLKQICSHLCLLRPGYSCACPOGSSFIEGTTTECDAAIELPIN 4376  
Qy 512 L-----R-----K 514  
Db 4377 LPPPCRMHGGNCYFDETDLPK 4398

RESULT 7

US-08-652-877-90

Sequence 90, Application US/08652877

Patent No. 6187548

GENERAL INFORMATION:

APPLICANT: Akerstrom, Goran

APPLICANT: Juhlin, Claes

APPLICANT: Rask, Lars

APPLICANT: Crumley, Gregg R.

APPLICANT: Morse, Clarence C.

APPLICANT: Murray, Edward M.

APPLICANT: Hjaln, Goran

TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments

TITLE OF INVENTION: Thereof and DNA Encoding Same

NUMBER OF SEQUENCES: 106

CORRESPONDENCE ADDRESS:

ADDRESSEE: Rhone-Poulenc Rorer Inc.

STREET: 500 Arcola Rd., 3C43

CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426-0107  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: Macintosh  
OPERATING SYSTEM: System 7.5.1  
SOFTWARE: Word 6.0 (Patentin)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/652,877  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/15203  
FILING DATE: 22-NOV-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/344,836  
FILING DATE: 23-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/487,314  
FILING DATE: 07-JUNE-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Savitzky, Martin  
REGISTRATION NUMBER: 29,699  
REFERENCE/DOCKET NUMBER: A1355E-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-454-3816  
TELEFAX: 610-454-3808  
INFORMATION FOR SEQ ID NO: 90:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4655 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-652-877-90

Query Match 70.3%; Score 2577.8; DB 3; Length 4655;

Best Local Similarity 9.8%; Pred. No. 4e-32;

Matches 431; Conservative 59; Mismatches 20; Indels 3892; Gaps 381;

Qy 1 M-----S-----I-----S-----S-SDE- 7  
Db 1 MDGPPAAVACTLLALVALAPASGOECDSAHFRCSGHCI PADWRC DGTGKDCSDDAEI 60  
Qy 8 ---VN---F-----L---V-----12  
Db 61 GCAVTCQGYFKCQSEGGCIPSSWVCDQDQDCDSDRQDCSQSTCSHQITCSNGQC 120  
Qy 13 ---VR-----Y---L-----Q-----E---S 19  
Db 121 IPSEVRCDHVRCDPGADENDCQYPTCEQLTCDNGACVNTSQKCDWKVCDRSDSINCT 180  
Qy 20 -----G-FS-----H-----S---A--F-T-----F---G-I-----30  
Db 181 EICLHNEFSCNGECIPRAYVCDHNDQDGSDEHACNYPTCGGYQFTCPSGRCIYQNWV 240  
Qy 31 -----ES-----H-----IS-----Q--SNI 39  
Db 241 CDGEDDCKNGDEDCESGPHDVHKSPREWSCPESGRCISYKVC DGLDCPREDEN- 299  
Qy 40 N--G-----AL-----VP-----PAALI-----S-----I--I-- 53  
Db 300 NTSTGYKSMTLCSALNCOYQCHETPYGGACFCPPGYIINHNSRVCVDFDQCIWICD 359  
Qy 54 QK-----G-L---QY-----VE 61  
Db 360 QKCESRPGRLHCHCEBEGYILERQYCKANDSFGEASII FSNGRDLLIGDIHGRSPRIIVE 419  
Qy 62 -----A-----E---VSI-----N 67  
Db 420 SQNRGVAVGVAFYHLQRVFWTDTVQNKVFSVDINGLN IQEVLNVSVETPENLAVDWNV 479

QY 68 ---E---DG---TL-----72  
Db 480 KIYLVETKVRIDWNLDGSRVTLITENLCHPRGIAVDTGYLFPFSDWESLSGPKLE 539  
QY 73 --F-DG--R-----P-----IE-----S 80  
Db 540 RAFMDGSRNRDLVKTKLWGPAGVTLDMSKRKYVWDSRFDYIETVTVYDGIQRKTVVHGS 599  
QY 81 L-----SL---I---D---AVM-----88  
Db 600 LIHPFGVSLFEGQVFTDWTAKVLANKANKFTETNPQVYQASLRPYGVTVYHSLRQPYA 659  
QY 89 --P--D-----V-V-----Q--T-----94  
Db 660 TNPCKNNGCEQVULSHRTDNDGLGRCKTFGQFQDTRHCTAVQNFIFSSQVAI 719  
QY 95 R-----Q-----Q-----97  
Db 720 RGIPFTLSTQEDVMVPVSGNPFVGDIDPAQDSTIFFSDMSKHMIFKQIDGTGREILA 779  
QY 98 -----A-----Y-----R--DK---LAQ---O---H--A--108  
Db 780 ANRVENVESLAFDWISKNLYWTDHSHYKSIISVMRLADKTRTVVQYLYNPRSVVHPFAGY 839  
QY 109 -----A-----A-----AA-----A-----114  
Db 840 LPFTDFRPKIMRAMSDGSHLLPVINTLWGNGLAIDWAASRLYWDAYDPKIEHSTF 899  
QY 115 -----A-----A-----A-----T-----118  
Db 900 DGLDRRLGHIEQWHPFLAIFGEHLFTDWRLGAIIRVKADGEMTVIRSGIAVILH 959  
QY 119 -----N--QGS--A-----KNG-----EN-----TA--N-----G---133  
Db 960 LKSYDNIOTGSNACNOPTHENGDCSHCFPFVFNQVRCGPGYMRLASNHLTCBGDPTN 1019  
QY 134 E--E-----N-----GA--H-----T-----140  
Db 1020 EPTEQCGLSFPCKNGRCVNPYYLDCGVDDCHDNDSEQLCGTLNNTCSSAFTCGHGE 1079  
QY 141 I--A-----N-----H-----T--DMME-----V--D-----GD---154  
Db 1080 IPAHWRCKDNKDCVDSDEHNCPTHAPASCLDTQYTCDNHQCSIKWVCDTNDNCGDSD 1139  
QY 155 -----VEI--PS-----N-----K--A-----162  
Db 1140 EKNCNSTETCOPSOFCNPNHRCIDLSPVCDGDKDCVDGDEVGCVLNCCTASQPKASGDK 1199  
QY 163 ---V-----VL-----R--G--H-----E-----169  
Db 1200 CIGVTNRCGVFDCSDNSDEAGCPTRPFGCHSDEFQCGEDGICIPNFWECGHPDCLYG 1259  
QY 170 S--E-----V-----F-----IC--AW-----N-----P-----179  
Db 1260 SDEHNACVPKTCPSSEYHCDNGNCIHRALCDRNDGCDMSDEKDCPTQPFRCPSQWQC 1319  
QY 180 -----V-----V-----SD-----182  
Db 1320 LGHNICVLSVVDGIFDCPNGTDESPLCNGNSCDFNGGCTHECVQEPFGAKCLPLGF 1379  
QY 183 LL-----V-----S-----GS-----189  
Db 1380 LLANDSKTCEDIDECIDLGSQHCYNWRGSRFCSDTGYMLSDGRTCKVTASESLLLL 1439  
QY 190 -----DS--T-----A-----RI--W-----N-----197  
Db 1440 VASQNKIADSVTSQVHNIYSLVENGSYIYAVDPDSISGRIFWSDATQGTWSAFQNGTD 1499  
QY 198 -----LSE-----N-----S-----T-----203  
Db 1500 RRVFDSIIILTETIADWGRNLYWTDYALETIEVSKIDGSHRTVLISKMLNPRGLAL 1559  
QY 204 -----S--G-----P---T-----207

Db 1560 DPMNEHLLFWSDGWHHPRIERASMDGSMRTVIVQDKIFWPCGLTIDYENRLLYFMDSYL 1619  
QY 208 -----Q-----Q-----LVLRH-----213  
Db 1620 DYMFCDYNGHRRQVIASDLIIRHPYALTLPEDSVYVTRATRRVMRANKWHGNGOSVV 1679  
QY 214 -----CI--R-----EG-----G-----219  
Db 1680 MYNIQWPLGIVAVHPKQPNVNPFCAFSCSHLLSSQPHFYSCVCSGWSLSPDLLN 1739  
QY 220 -----Q-----D--VP--S---N---K--D-----227  
Db 1740 CLRDDQPLITVRQHIIFGILNPEVKSNDAMVPIAGIQNGLDVEPDAAEQIYVWENPG 1799  
QY 228 --V-----T-----S---LDW-----N---S--E---G---T 238  
Db 1800 EHRVKTGDTNRTVFASISMVGPSMLALDWISRLNLYSTNPRTQSTEVLTGLHGDIRYKT 1859  
QY 239 LLA--T-----G--S---Y-----DG-----F-----248  
Db 1860 LIANDGTALGVGPPIGITVDPAKGKLYWSDQGTDSGVPAKIASANMDGTSVKTLFTGNLE 1919  
QY 249 -----A-----R--I-----W-----T-----253  
Db 1920 HLECVLTDIEQKLYWAVTGRGVIERGNVDGTRMILVHQLSHPMGIAVHDSFLYTTDEQ 1979  
QY 254 --K--D--G--N---L---ASTL--G--O--H-----K-----266  
Db 1980 YEVIERRKATGANKIVLRDNPVNLRLGLQVYHRRNAESNGCSNNMACQQLCLPVP 2039  
QY 267 -----G-----P-----I--FALK-----273  
Db 2040 LFSCACATGKLPNDNRSCSPNSFIVSMLSAIRGFSLELSDHSETMVPVACQGRNALH 2099  
QY 274 -----W-----N-----K-----K-----GN--279  
Db 2100 VDVVSSGTYWCDFFSSVASDANIRRIKPDGSSLMNIVTHGIGENGVRGIAVDWAGNL 2159  
QY 280 -----F-----F-----I-----281  
Db 2160 YFTNAFVSETLIEVLINTYRRVLLKVTVDMPRHIVDPKPNRYLFWADYQRPKIERSP 2219  
QY 282 --L--S-----A-----G---VD-----287  
Db 2220 LDCNTEVLVSGIVTPRGVLAVDSDGYVWVDDSLDIARIRINGENSEVIRYGSRYPT 2279  
QY 288 -----K-----T-----T--I---I--W---D-----294  
Db 2280 PYGITVFENSIWVDRNLKKIQAOSKEPENTEPPTVIRDININWLRDVTIPDKVQPRSPA 2339  
QY 295 -----A-----HT-----G-----E-----A-----300  
Db 2340 EVNNPCLENNGCCHLCPALPGLHTPKDCAPGTQSDGKNCAISTENFLIFALSNSLR 2399  
QY 301 -----K-----Q-----O--F-----P-----305  
Db 2400 SLHLDPENHSPPTQINVERTYMSLDYSDSDRIYFTQNLASGVGQISYATLSSGIHTPT 2459  
QY 306 -----F-----306  
Db 2460 VIASGIGTADGIAFDWITRRIYVSDYLNQMINSMAGDSNRTVIARVPKPRAILVDP 2519  
QY 307 -----H-----S--AP---ALD-----313  
Db 2520 YLWAWDWDHAKIATERATLGNFRVPIVNSGLVWPSGLTDYEEDLLYWDASLQRIERST 2579  
QY 314 ---VD-----W-----Q-----S-----N-----319  
Db 2580 LTGVDRVIVNAAVHAFGLTYGYIYWTDLYTORIVRANKYDGSQGIAMTLLSQPRG 2639  
QY 320 -NT-----F-----322

Db 2640 INTVKNKQOCCNPPCEQFNGGSHICAPGNGAECQCPHEGNYLANNRKHICVDNGER 2699  
Qy 323 --AS--CS-----TD-M--C-IH-----V----- 333  
Db 2700 CGASSSTCSNGRCISBEWKCDNDGDSDESVCALHTCSPTAFTCANGRCVQYSR 2759  
Qy 334 CKL-----G--QD-----R-P--I----- 342  
Db 2760 CDYNDGDSDEAGLFRDCNATTFMCMNRRCIPREFICNGVNDCHDNTSDEKNC 2819  
Qy 343 -----K-----T-----FQ----- 346  
Db 2820 RTCQGYTKCHNSNICIPRYLDCGNDGDSNENPTYCTHTCSSEFQCASGRICPQ 2879  
Qy 347 -----GH--T----- 349  
Db 2880 HWYCDQETDCFDASBPASGHSERTCLADEFKDCGRCIPSEWICGDNDCGMSDEK 2939  
Qy 350 -----N--E--VNA-----IK-W----- 357  
Db 2940 RHQONQNCSDSEFLCVDNRPDRRCIPQSWVCDGVDCTGDYDENQNCNTRTCSNEFT 2999  
Qy 358 ----- 357  
Db 3000 CGYGLCIPKIFCRDRHNDGDSYDERGCLYQTCQNOFTCQNGRCISKTFVCDENDCGD 3059  
Qy 358 --D-----PT-----G-----N--L----- 363  
Db 3060 GSDLMHLCHTPEPTCPPEHFKDNGRCIEMMKLNHLDCLDSDGKGGINECHDPSI 3119  
Qy 364 -----LAS--CS-----D-D--M----- 371  
Db 3120 SGCDHNCTDLTFSFYCSRPGYKLMDSKRCVDECTEMPFVCSQKNCENVISYICKCA 3179  
Qy 372 -----T-----L----- 374  
Db 3180 PGYLREPDKTCQNSNIEPYLFSNRYLRNLTIDGYFYSLEGLDNVVALDFRVEK 3239  
Qy 375 -----I-----W-----S-- 377  
Db 3240 RLYWIDTQVIERMFLNKETIINHRLPAESLAVDWVSRLYLDARLDGLFVSDL 3299  
Qy 378 -----M-KQ--D--N--C-----VHDLO--A--H-----NK-- 392  
Db 3300 NGHRRMLAQHCVDANNTFCFNDPRGLALHP-QGYLYWADWGHRAVIGRVGMDGNTKSV 3358  
Qy 393 -----E-----I-YT-----IK----- 398  
Db 3359 IISTKLEWPNGITIDYNDLLYWADAHGLGYEYSDLEGHHRHTVYDGLPFPFAITIPED 3418  
Qy 399 -----W-----S-P--T--G-- 403  
Db 3419 TIYWTWNTRTVEKNGKNDGSRQTLNTHTRFPDIHVTHPYQPIVSNPCGTNNGGCSH 3478  
Qy 404 -----PG-----T-----NN----- 408  
Db 3479 LCILIKGGKGTCECPDFTLQSGSTYCMPCSSSTQFLCANNEKCIPIWKKCDQKDC 3538  
Qy 409 -----P-----NA--N-----LML-----AS- 417  
Db 3539 SDGSDELALCPQRCRLGQFCSDGNCSTPOTLCNAHQNCPDGSDBDRL-LCENHHDCSN 3597  
Qy 418 -----A-----S-----F-----DS-----T-----VR-----L--W--DV- 429  
Db 3598 EWOCAKRCIPESWQCDTNDCEDESDSSHCASRTCPGQRCANGRCIPQAWKCDVD 3657  
Qy 430 -DRG-----I--CI--H-----T-L----- 438  
Db 3658 NDCGDSDEPIBECMSSAHLCDNFTEFSCKTNRYCIPKWAQVNGVDDCRDSDSQCEER 3717  
Qy 439 T-----K-H-----Q-----E--P-----V----- 445  
Db 3718 TCHPVGDFRCKNHHCIPLRWQCDGQNDGDSNDEENCAPRECTESEFRVCNQCIPSRWI 3777

Qy 446 -----Y--S--V-----A-----FS 451  
Db 3778 CDHYNDGDSNDEROCMETCHPEYFQCTSGHCVHSELKCDGSDASDCDASDEADCTRF- 3836  
Qy 452 PDGRYL-AS-----GS-----F----- 462  
Db 3837 PDGAYCOATMFCCKNHVCIPPYWKCDGDDDCGDSDEELHCLDVPFCNSPNRFRCDNNRC 3896  
Qy 463 -----D-----K-----CV-H--I-----W 469  
Db 3897 IYSHEVNGVDCGDTDETEHCRKPTPKPCTEYKCGNGHCIPHDNVCDADDGCDW 3956  
Qy 470 -----N-----T-----Q--T-----G-----A-----L-V----- 477  
Db 3957 SDELGCNKKERTCAENICEQNCQTQNEGFFICCTAGTETNVDFRTSCLDINECBQFGT 4016  
Qy 478 --H-----SY-----R----- 481  
Db 4017 CPQHCRNTRKGSYECVACDGFTSMRDRPKRCAAGSSPLLLLPDNRVIRKYNLSRFRSE 4076  
Qy 482 -----G-----T-----G--G--G--I--FE-----V-C-- 490  
Db 4077 YLQDEBYIOAVDYDNDPEDIGLSVVYTVRGSGRFGAIKRAYIPNFESGRNNLVQEVDL 4136  
Qy 491 -----W-----N-----AA----- 494  
Db 4137 KLYWMPDGIADVWGRHIYMSDVKNKRIEVAKLDRYRKWLISDLDQPAIAVNPKL 4196  
Qy 495 -----G-----D-K--V----- 498  
Db 4197 GLMFWTDMGKEPKLESAMNGEDRNILVPEDLGWPTGLSIDVLLNDRIYMSDFKEDVIET 4256  
Qy 499 -----G-----A-SA--S-----D-----G----- 505  
Db 4257 IKYDGTDRVIAKEAWNPYSLDIFEDQLYWSKEGVEWQKNGKQKKEKTLVNPWLT 4316  
Qy 506 -----SV--C--V-----L-----D----- 511  
Db 4317 QVRIHQRLYKNSVNPMLCKQICSHLCLLRPGGYSACPOGSSFIEGTTTECDAAIELPIN 4376  
Qy 512 L-----R-----K 514  
Db 4377 LPPPCRCMHGNCYFDETDLPK 4398

## RESULT 8

US-08-652-877-86  
; Sequence 86, Application US/08652877  
; Patent No. 6187548

## GENERAL INFORMATION:

; APPLICANT: Akerstrom, Goran  
; APPLICANT: Juhlin, Claes  
; APPLICANT: Raek, Lars  
; APPLICANT: Crumley, Gregg R.  
; APPLICANT: Morse, Clarence C.  
; APPLICANT: Murray, Edward M.  
; APPLICANT: Hjalms, Goran  
; TITLE OF INVENTION: Human Calcium Sensor Protein, Fragments  
; TITLE OF INVENTION: Thereof and DNA Encoding Same  
; NUMBER OF SEQUENCES: 106  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Rhone-Poulenc Rorer Inc.  
; STREET: 500 Arcola Rd., 3C43  
; CITY: Collegeville  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19426-0107  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: Macintosh  
; OPERATING SYSTEM: System 7.5.1  
; SOFTWARE: Word 6.0 (Patentin)

Qy	1	M	-----S-----I-----S-SDE-7
Db	1	M	RGPAAVACTLLIALVACLAPASQECDSAHRGCGSHCIPADWRCDGTKDCSDDAEI 60
Qy	8	---VN---	F-----L-----V-----12
Db	61	G	CAVYTCQGYFKOSEGOCIFSSVWCDDQDCDGSBERQCSQSTSSHQITCSNGQC 120
Qy	13	---YR---	Y-----L-----Q-----E--S 19
Db	121	I	PSEYRCDHVRDCPDGADENDCQYPTCQLTCDNGACYNTSQKDWKVDCHDSDSEINCT 180
Qy	20	---G-FS---	H-----S---A-F-T---P---G-I---30
Db	181	E	ICLHNEFSCNGECIPRAYVCHDNDQDQGSDEHACNPTCGGYQFTCPSGRCIYQNWV 240
Qy	31	---ES---	H-----IS-----Q--SNI 39
Db	241	C	DGEDDKNGDEDCGSPHVDHKCSPREWSPESGRCSIIYKVCQDILDCPREN- 299
Qy	40	N--G---	AL-----VP-----PAALI-----S-----I-I--53
Db	300	N	TSTGYCSMTLCSALNCQYQCHETPYGCACFCPPGYIINHDSRTCVEFPDDCQIGICD 359
Qy	54	QK-----	G-L--QY-----VE 61
Db	360	Q	CESRPGRHLCHCEGYILRGQYCKANDSFGESAIIFNSGRDLIDIGHRSFRILVE 419
Qy	62	---A---	A-----E--VSI-----N 67
Db	420	S	QNGVAVGVAHYHLQRVFWTDTVQNKVFSVDINGLNIQEVLVSVETPENLAVDWVN 479
Qy	68	---E---	DG-----TL-----72
Db	480	K	IYLVETKNRIMDWNLDGSRVLTITENLGHPRGIAVDPTGYLFFSDWESLSGEKLE 539
Qy	73	--F-DG--R-	-----P-----IE-----S 80
Db	540	R	AFMDGSRKOLKVTKLGPAGVTLDMISKRVYVWDSRFDYIETVTDGIQRKTVVHGGS 599
Qy	81	L-----	SL---I---D---AVW-----88



Db 1680 MYNIQWPLGIVAHVHPKQPNVNCAPFRCRSHLCLJSSQGHFHYVCVPCSGWSLSPDLN 1739  
Qy 220 -----Q-----D--VP-S--N--K-D-----227  
Db 1740 CLRDDQPLITVRQHIIIFGILSNPEKSNDAWPIAGIQNGLDVEPDDAEQYIYWENPG 1799  
Qy 228 -----V-----T-----S-----LDW-----N-----S-E-----G-----T 238  
Db 1800 EHRVKTGTRNTVFASISMGVSPMNLALDWISRLNSTYNTPTQSIIEVLTLHGDIRYKT 1859  
Qy 239 LLA--T-----G-S-----Y-----P-----DG-----F-----248  
Db 1860 LIANDGTALGVGPPIGTVDPARKLYWSDQGTDSGVPAKIASANMDGTSVKTLFGNLE 1919  
Qy 249 -----A-----R-I-----W-----T-----253  
Db 1920 HLECVTLIDIEQKLYWAVTGRGVIERNVDGTRMILVHQLSHPWGIAVHDSFLYTDEQ 1979  
Qy 254 -----K-D-----L-----ASTL-G-Q-H-----K-----266  
Db 1980 YEVIERVKATGANKIVLRDNVNLRLGLQVYHRRNAEASSNGCSNNNNAQQICLPVPGG 2039  
Qy 267 -----G-----P-----I-----PALK-----273  
Db 2040 LFSCACATGKLPNDNRSCSPYNSFIVVMSLAIIRGFSLSLSDHSETWVPVAGQGRNALH 2099  
Qy 274 -----W-----N-----K-----K-----GN- 279  
Db 2100 VDVVSSGFIYWCDFSSVASDNAIRIKPDGSLMNIIVTHGIGENGVRGIAVDWAGNL 2159  
Qy 280 -----L-S-----A-----G-----VD-----I-----281  
Db 2160 YFTNAFVSETLIEVLINTTYRRVLLKVTVDMPRHIVVDPKRYLFWADYGQPKIERSF 2219  
Qy 282 -----L-S-----A-----G-----VD-----287  
Db 2220 LDCTNRTVLVSEGIPTVPRGLAVDRSDGYVYVWDSLDIIAIRINGENSEVIRYGRYPT 2279  
Qy 288 -----K-----T-----T-I-----I-W-----D-----294  
Db 2280 PYGITVPENSIWDRNLKXIFOASKEPENTPEPTVIRDNINWLRDVTDFKOVQPRSPA 2339  
Qy 295 -----A-----HT-----G-----E-----A-----300  
Db 2340 EVNNNPCLENGGCHLCPALPGLHTPKDCARGLQSDGKNCAISTENFLIIFALSNSLR 2399  
Qy 301 -----K-----Q-----Q-----F-----P-----305  
Db 2400 SLHDPENHSPPQTINVERTVMSLDYSVSDRIYFTQNLASGVQISYATLSSGIHTPT 2459  
Qy 306 -----F-----P-----Q-----F-----306  
Db 2460 VIASGIGTAGIADPWITRIIYSDYLNQMINSAEDGSRNRTVIARVPKPRAILVDPCCG 2519  
Qy 307 -----H-----S-----AP-----ALD-----313  
Db 2520 YLYWADWDHAKIERATLGNFRVPIVNSLVPSGLTLDYEEDLLYWWDASLQRIERST 2579  
Qy 314 -----VD-----W-----Q-----S-----N-----319  
Db 2580 LTGVREVIYVAAVHAFGLTYGYIYWTDLTYRIYRANKYDGSGLIAMTTNLLSQPRG 2639  
Qy 320 -----NT-----F-----322  
Db 2640 INTVVRNQKQCCNPNCEQFNGGSHICAPGNGAECQCPHEGNWYLANNRKHICVDNGER 2699  
Qy 323 -----AS-----CS-----TD-M-----C-IH-----V-----333  
Db 2700 CGASSFTCSNGRCISEBWKCDNDNDGSDGDEMSVCALHTCSPTAFTANGRCVQYSYR 2759  
Qy 334 CKL-----G-----QD-----R-----P-----I-----342  
Db 2760 CDYNDGSDGSDRAGCLFRDNCATTEPMCNRRRCIPREFICNGVDNCHDNNNTSDEKNCPD 2819

Qy 343 -----K-----T-----FQ-----346  
Db 2820 RTCQSGYTKCHNSNICIPRYLDCGNDNCGNSDENPTYCTTHTCSSSEFQCTSGRCIPQ 2879  
Qy 347 -----GH-----T-----349  
Db 2880 HWYCDQETCFDASDPASCGHSERTCLADBFKCGGRICIPSEWICDGDNDGMSDEDK 2939  
Qy 350 -----N-----E-----VNA-----IK-----W-----357  
Db 2940 RHOCQONQSDSEFLCVDNRPDRRCIPQSWVCDGDVDTGDYDENQONCTRRCTSENEFT 2999  
Qy 358 -----357  
Db 3000 CGYGLCIPKIFRCDRHDCGYSYDERGCLYQTCQNOFTCQNGRCISKTFVCDENDCGD 3059  
Qy 358 -----D-----PT-----G-----N-----L-----363  
Db 3060 GSDELMHLCHTPEPTCPPEHFKCDNGRCIEMMKLCHLDDCLDCLNSDEKGCINECHDPSI 3119  
Qy 364 -----LAS-----CS-----D-----D-----M-----371  
Db 3120 SGCDHCTDTLTSFYCSCRPGYKMSDKRTCVDIDECTEMPFVCSOKCENVISYICKCA 3179  
Qy 372 -----T-----L-----K-----374  
Db 3180 PGYLRBPDKGTCEQNSNIEPYLIFSNRYLRNLITIDGYFYSLLILEGLDNNVALDFORVEK 3239  
Qy 375 -----I-----W-----S-----377  
Db 3240 RLYWIDTQVIERMELFNKTNETIINHRLPAEASLAVDWVSRKLYWLADRLDGLFVSDL 3299  
Qy 378 -----M-----KQ-----D-----N-----C-----VHDLO-----A-----H-----NK-----392  
Db 3300 NGHREMLAQHCVDANNTFCFNPRGLAHP-QYGYLYWADWGHRAYIGRVMDGTNKS 3358  
Qy 393 -----E-----I-----YT-----IK-----398  
Db 3359 IISTKLEWPNGITIDYTNDLLYWADAHGLYIEVSDLEGHHRHTVYDGLPHPAITIFED 3418  
Qy 399 -----W-----S-----P-----T-----G-----403  
Db 3419 TIYWTDMNTRTVEKGNKYDGSNRQTLVNTTHRPDIHVHYPRQPIVSNPCGTNNGCSH 3478  
Qy 404 -----PG-----T-----NN-----408  
Db 3479 LCLIKFGKGFTCECPDPFRTLQLSGTYCMPMCSSTQFLCANNEKCIPIWVKDQKOC 3538  
Qy 409 -----P-----NA-----N-----LML-----AS-----417  
Db 3539 SDGSDELALCPQBFCELGQFCSDGNGCTSPOTLCNAHQNCPOGSDERDL-LCENHHCDN 3597  
Qy 418 -----A-----S-----F-----DS-----T-----VR-----L-----W-----DV-----429  
Db 3598 EMQANKRCIPESWQCDTENDCEDNDESDSHCASRTCPGQFCANGRCIPQAWKCDVD 3657  
Qy 430 -----DRG-----I-----CI-----H-----T-----L-----438  
Db 3658 NDCGDHSDPEIBECMSAHLCDNFTEFCKTNRNRCIPKWAVCNVDDCDRNSDEQCEER 3717  
Qy 439 -----K-----H-----Q-----E-----P-----V-----445  
Db 3718 TCHVGDGFRCKNHHCIPLRWQCDGQNDGNDSENCAPRECTESBFRVCNQOCIPSRWI 3777  
Qy 446 -----Y-----S-----V-----A-----FS-----451  
Db 3778 CDHYNDGNSDERDCEMTCHPEYQCTSGHCVHSELKCDGSADCLDASDEADCPTRF- 3836  
Qy 452 PDGRYL-AS-----GS-----P-----462  
Db 3837 PDGAYQATMFECKNHVCIPIPYWKCDGDDCGDGSDEELHLCLDVPNCSPNRRFCDDNRC 3896

QY 463 -----D-----K-----CV-H-I-----W 469  
Db 3897 IYSHEVNGVDDCGDGTDETEHCRKPTPKCTEYKYKNGCHPHDNVNCDDADDGCDW 3956  
QY 470 -----N-----T-----Q-T-----G-----A-----L-V----- 477  
Db 3957 SDELGNCKGKERTCAENICQNCQTLNEGFGICCTAGFETNVFDRTSCLDINECQFGT 4016  
QY 478 -----H-----SY-----R----- 481  
Db 4017 CPQHCRTKGSYECVACADGFTSMRPGKRCAAEGSSPLLLLPDNRKRYNLSRFSF 4076  
QY 482 -----G-----T-----G-----G-----I-PE-----V-C----- 490  
Db 4077 YLQDEEYQAVDWDMDXDGSLSVVYTVRGEGRFGAIKRAIYIPFESGRNNLVQEVDL 4136  
QY 491 -----W-----N-----AA----- 494  
Db 4137 KLKVMQPDGIADVWGRHIYWSVDKKNRIEVAKLDRYKWLJSTDLQDPAIAVNPKL 4196  
QY 495 -----G-----D-K-V----- 498  
Db 4197 GLMFWDGKPKXESAMNGEDRNILVFDLWPTGLSIDYLNDRYWSDFKEDVIET 4256  
QY 499 -----G-----A-SA-----S-----D-----G----- 505  
Db 4257 IKYDGTDRRVIKXAMNPYSLDIFEDQLYWISKEGEVWQKNGFGQKKEKTLVNPWLT 4316  
QY 506 -----SV-----C-V-----L-----D----- 511  
Db 4317 QVRIFHQLRYNKSVPNLCKQICSHLCLLRPGGYSCACPGSSFIEGSTECDAAIELPIN 4376  
QY 512 L-----R-----K 514  
Db 4377 LPPPCRMHGNCYFDETDLPK 4398

## RESULT 9

US-08-718-388-9  
; Sequence 9, Application US/08718388  
; Patent No. 6271362  
; GENERAL INFORMATION:  
; APPLICANT: MORIKAWA, MINORU  
; APPLICANT: HARADA, NAOKI  
; TITLE OF INVENTION: GENE ENCODING I9G Pc REGION-BINDING  
; TITLE OF INVENTION: PROTEIN  
; NUMBER OF SEQUENCES: 29  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: BIRCH, STEWART, KOLASCH AND BIRCH  
; STREET: PO BOX 747  
; CITY: FALLS CHURCH  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22040-0747  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/718,388  
; FILING DATE:  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MURPHY JR, GERALD M  
; REGISTRATION NUMBER: 28,977  
; REFERENCE/DOCKET NUMBER: 0230-111  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 205-8000  
; TELEFAX: (703) 205-8050  
; INFORMATION FOR SEQ ID NO: 9:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 5405 amino acids

; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-718-388-9  
Query Match 69.8%; Score 2560.4; DB 3; Length 5405;  
Best Local Similarity 9.6%; Pred. No. 2e-31;  
Matches 410; Conservative 81; Mismatches 17; Indels 3766; Gaps 353;  
QY 1 M-----S-I-----SS-D-EVN-----FLV-----YR-----Y----- 15  
Db 1 MGALWSWILWAGATLLWGLTQEASVDLK-NTGREBEFLTAFLQNYLAYSKAYPRLLISS 59  
QY 16 -----L-Q-----ES-----G-F-S-H----- 23  
Db 60 LSESPASVILSQADNTSKKTVRPGESVMVNISAKAEMIGSKIFOHAVVIHSDYALSIVQ 119  
QY 24 -----SA-----F-----T 27  
Db 120 ALNAKPDPTAELTLRPIQALGTEYFVLTPPGTSARNVKFAVVAGAAGASVSLKGSVT 179  
QY 28 F-----G-----I 30  
Db 180 FNGKFYPAGDVLKRVTLQPNVAQLQSSVDLSGSKVTASSPVAVLGSHSCAQKHTTCNHVV 239  
QY 31 E-----S-H-I-----SOS-----N-I----- 39  
Db 240 EQLLPTSAWGHYVVTPLASQSYDLAFVVAQATKTYNHGGITGSRGLQAGDVVEFEV 299  
QY 40 -----N-----GAL-----V-P-----P-AAL-----I-S----- 51  
Db 300 RPSWPLYLSANVGIVLLFGTGAIKNEVTVDPVLVLPDVAAYCPAVVVKSVPGCEGVAL 359  
QY 52 II-Q-K-----G-----IQY-V-----E-AEV-----S-I-----N----- 67  
Db 360 VVAQTKAISGLTIDGHAVGAKLTWEAVPGSEFSYAELGTADMIHTABATNLGLLTG 419  
QY 68 --E-----D-G-T-L----- 72  
Db 420 LAKAIGYATAADCGRTVLSFVPEPSCEGMAQACQRCQVVGKGAGCVAESTAVCRAQGDPH 479  
QY 73 ---FDGR-----P---IE----- 79  
Db 480 YTTFDGRRYDMGTCSTYTWELCSEDDTLPAFSVEAKNEHRSRRVSVVGLVTVRAYSHS 539  
QY 80 -SLS-----LID-----A-V-----M-----P-----D-----VW-----Q----- 93  
Db 540 VSLTRGEVGFVLVDNQSRPLVSLSEGLRVYQSGPRAVVELVFLVTVTYDWDQCLALS 599  
QY 94 -TR-Q-QA-----Y-----RD-----KL-----A 104  
Db 600 PARFQDVCGLCGNYNGDPADFLTPDGLAPDAVEFASWKLDDGDLCEGCGQNNCPA 659  
QY 105 ---Q-QH-----A----- 108  
Db 660 CTGQAQHYEGEDRLCGMLTKLDGPFVAVCHDTLDPFLEQCYYDLCVVGGERLSLGRGLS 719  
QY 109 A-A-A-----A-----A-----AAA-TN----- 119  
Db 720 AVAQACLELGISVGDWRSPANCPLSCPANSRYELCGPACPTSCNGAAAPSNCSGRPCVEG 779  
QY 120 -----Q-Q-----G-S-A-K-----NGEN-----T----- 130  
Db 780 CVCLPGFVASGGACVPASCGCTFGQLQAPQEVWVADELCCRRCTCNGATHQVTRCDKQ 839  
QY 131 ---ANGE-----EN-----GA-H-----T-----I-----AN----- 143  
Db 840 SCFA-GERCQVQNGLLGCGYDPDRFGTCQSGDPHYVVSFGDGRFPDFMGTCVTVLLVSGSCQNA 898  
QY 144 -----N-H-----T-----DMME----- 150  
Db 899 ALPAFRVLVENEHRGSGTYSYTRAVRVEARGVKVAVRVEYPGQVLDVDDVLQYLPFOAADG 958

QY 151 -----V-D-G-----D-----V-EIPS----- 159  
Db 959 QVQVFRQRODAVVRTDFGLTVTYDMNARVTAKEPSSVAEALCGLGNFNGDPADDLALRG 1018  
QY 160 -----K-K-G-NF-----I-----LSA-----G 285  
Db 1019 GGOAANALAFNGWQEBTRPGCGATEPGDKLDSLVAQLOSKNCGILADKPPREC 1078  
QY 162 -----AV-----VL-R-----G-H----- 168  
Db 1079 HSKLDPOGAVDVCYDECLLPQSGPLCDALATYAAQAGATVHPWRSEELCPISCPP 1138  
QY 169 -----ES----- 170  
Db 1139 HSHYACSYCLPLSCGDLPPVGGCGSECHGCVCEGDFALSGESCLPLASCGCVHQTGTH 1198  
QY 171 -----E-----V----- 172  
Db 1199 PPGQTFYPGCGDSLCHQCBGLVSCSSCGPHEACQPSGSLGCVAVGSSCQASGDP 1258  
QY 173 -----F-----I-----C-----AW----- 177  
Db 1259 HYTFDGRDFMGTCVYVLAQTCGTRPGLHRFAVLQENVANGRVSUTRVTITQVANF 1318  
QY 178 -----N-----PV-----SD-----L-----VS- 186  
Db 1319 TLRLEQRMKVTVNGVDMKLPVVLANGQIRASQSGDVVITETDFGLRVAYDLVYVVRTV 1378  
QY 187 -----G-SGD-----S-TARI-----W----- 196  
Db 1379 PGNYQOMCGLGNYGDPKDFQKPNGSQAGNA-NEFGNSWEEVVPDSFCLPPTCPFG 1437  
QY 197 ----- 197  
Db 1438 SEDCIPSHKCPPELEKYYQKEEFCGLLSPTGFLSSCHKLVDPPQPLKDCIFDLCLGGN 1497  
QY 198 LS-----E-NS-----T----- 203  
Db 1498 LSLCSNIHAYVSACQAAGHVEPWRTEPCPMECPNPNHSHYELCADCATCSLGCALSAPQ 1557  
QY 204 -----SG-----PTQ-----L-----VL-----R----- 212  
Db 1558 QDGCBAEGCQDSGLYNGQACVPIQCCGCHNGVYYPEQTVLIDNCRQOCTCHAGKM 1617  
QY 213 -----H-----C-----I-----R-----E-GGO----- 220  
Db 1618 VCOEHSCKPGQVCPQSGGILSCVTKDPCHGVTCRPOETCKEQQGGQVCLPNEYATCWLWG 1677  
QY 221 -----D----- 224  
Db 1678 DPHYSFDGRKDFQGTCTNYLATTGCPGVSTQGLTPFTVTTKNQNRGNPAVSVYRVTV 1737  
QY 225 -----N-----KD-----VT-----SLDWN- 234  
Db 1738 AALGTNISIHKDEIGKRVNGVLTALPVSVADGRISVTOGASKALLVADFGQLQVSYDMW 1797  
QY 235 -----S----- 236  
Db 1798 RVDVTLPSSYHGAVCLGMDRNPNDQVFPNGTLAPSIPIWGGSWRAPGWDPLCWDDEC 1857  
QY 237 -G--T-----L-----LA--TG----- 243  
Db 1858 RGSCTCPEDRLBQYEGPGFGLAPGTGPGFTTCHAHVPPESFFKPGCVLDVCMGGGDRD 1917  
QY 244 -----SY----- 245  
Db 1918 ILCKALASYAACQAAGVIEDWRAQVCEITCPENSHYEVCGPPCPASCPSAPLUTTEA 1977  
QY 246 -----D-GF--A-R-I-----W--T-----K--DGN----- 257  
Db 1978 VCEGPCVEGCQCDAGFVLSADRCPLNNGCGCWANGTYHAGSEFWADGTCSQWCRCPG 2037  
QY 258 ---L---AS-----TL---GOHKG--PIF-ALK--W-----N----- 275

Db 2038 GGSLVCTPASCGLGEVCGLLPSGQH-GCQPVSTA-ECQAWGDHPHYVTLGDHRENFQGTCE 2095  
QY 276 -----KK-G-NF-----I-----LSA-----G 285  
Db 2096 YLLSAPCHGPPLGAENFTVTVANEHRGSAVSTRVTLQIYNHSLTLSARWPKLQVDG 2155  
QY 286 V-----DK-----TT----- 290  
Db 2156 VFVTLFPQLDSLHLHLSGADVVTITTSGLSLAFDGDSPVLRVPAAYAGSLGCLGNYN 2215  
QY 291 ----- 292  
Db 2216 QDPADDLKAVGKPAQWQVGGAGCGECVSKPCPSPCTPEQBSFGPDACGVISATDGP 2275  
QY 293 -----W--D----- 294  
Db 2276 LAPCHGLVPPAQYFQGCCLLDACQVQHGGCLCPAVATYAAQAGAQURWRPDCPPF 2335  
QY 295 ---AHT-----G-E-A-K-----Q----- 302  
Db 2336 QCPAHSHYELCGDSCPGSCPSLSAPEGCSACREBGCVCDAFVLSGDTCTVPVCGCCLHD 2395  
QY 303 -----Q-F----- 304  
Db 2396 DRYYPGLQTFYFPGPGCDLCLCRREGGEVSCPESSCGPHETCRPSGSLGCVAVGSTTQCA 2455  
QY 305 ---P-----F-----H-----S----- 308  
Db 2456 SGDPHYTFDGRDFMGTCVYVLAQTCGTRPGLHRFAVLQENVANGRVSUTRVTITQ 2515  
QY 309 -A-----P----- 313  
Db 2516 VANFTLRLEQRMKVTVNGVDMKLPVVLANGQIRASQSGDVVITETDFGLRVAYDLVYV 2575  
QY 314 ---V-----D-----W----- 316  
Db 2576 RVTVPNGYQLMCGLCGNYGDPKDDFKPNGSQAGNANEFNGNSWEEVVPDSFCLPPTC 2635  
QY 317 -----Q----- 317  
Db 2636 PPSBEGCIPSEPCPELEKYYQKEEFCGLLSPTGFLSSCHKLVDPPQPLKDCIFDLCLG 2695  
QY 318 -----SN-----N-TF-----A-S-CS--T- 327  
Db 2696 GGNLSILCSNIHAYVSACQAAGHVEPWRNETFCPMECPNPNHSHYELCADCATCSLGCALS 2755  
QY 328 -----D-----MC-I-----H-----V-----C--- 334  
Db 2756 PLQCPDCAEGCQDSGLYNGQACVPIQCCGCHNGVYYPEQTVLIDNCRQOCTCHAG 2815  
QY 335 KL-----GO-----D-----RP-----I----- 342  
Db 2816 KYVVOEHSCKPGQVCPQSGGILSCVTKDPCHGVTCRPOETCKEQQGGQVCLPNEYBATCW 2875  
QY 343 -----K-----T-F-----Q----- 346  
Db 2876 LMGDPHYHSPDKRKFQGTCTNYLATTGCPGVSTQGLTPFTVTTKNQNRGNPAVSVYRV 2935  
QY 347 -----G-----HTNE-----VN-----A-I-----K----- 356  
Db 2936 VTVAALGTNISIHKDEIGKRVNGVLTALPVSVADGRISVAQASKALLVADFGQLQVSYD 2995  
QY 357 --W--D--PT-----GNL-----LA-----S----- 366  
Db 2996 MNWRVDVTLPSYHGAVCLGCLGMDRNPNDQVFPNGTLAPSIPIWGGSWRAPGWDPLCW 3055  
QY 367 --C--S-----D----- 371  
Db 3056 DECRGSCPTCPEDRLBQYEGPGFGLAPGTGPGFTTCHAHVPPESFFKPGCVLDVCMGGG 3115  
QY 372 ---TL-K-----I--W-----SM----- 378

Db 3116 DHDILCKALASYAAQAGVWIEDWRAQVCEITCPENSHYEVCGPPCASPAPLPT 3175  
QY 379 -----K-Q-----D-----N----- 382  
Db 3176 TPAVCEGPCVEGCQDAGFVLSADRCVPLNNGCGCWANGTYHEAGSEFWADGTCSQWCR 3235  
QY 383 -----C-----V-----H-DLQ--- 388  
Db 3236 GPGGSLVCTPASGLGEVCGLLPSGQHGCPQVSTAEQAWGDPHYVTLGHRFDQGT 3295  
QY 389 -----A-H-----N-----I-YT-----KW----- 399  
Db 3296 EYLLSAPCHPPLCAENFTVTVAHEHRSQAVSYTRSVTLQIYNHSLTLSARWPKLQVD 3355  
QY 400 -----S-P-----T-----G-----P-----GT-----N- 407  
Db 3356 GVFVTLFPFQDLSLHAHLSGADVVTVTSLSLAFDGDVFLRVPAAYAGSLCGLGNY 3415  
QY 408 N-PNA-NLM-----L-----A-----SAS- 419  
Db 3416 NQDP-ADDLKAVGGKPAQVQGAQGCCEVCSPCTPEQOESEFGGPDACGVISATD 3474  
QY 420 -----F-----D-----ST-V-----RLW----- 427  
Db 3475 GPLAPCHGLVPPAQYFOGCLLDACQVQHGPGCLCPAVATYVAAQQAQALREWRPDCF 3534  
QY 428 -----D-----V----- 429  
Db 3535 PFQCPAHSHVELCGDSCPGSCPSLSAPEGESACRECVCDAGFVLSGDTVCPVQCGCL 3594  
QY 430 --DR-----G-I-----CI----- 435  
Db 3595 HDBRYPLGQTFYPGPGCDLRCRGEGEVSCBPSGCPHETCRPSGSLGCVAVGSTTC 3654  
QY 436 -----H-----T-----L-----TK-----H-----QE----- 443  
Db 3655 QASGDPHYTFDGRFRFDMGTCTVYVLAQTCTGRLHFAVLQENVAWNGRVSVRVIT 3714  
QY 444 -----PV----- 445  
Db 3715 VQVANFTLRLEQRQKVTYVNGVDMKLPVVLANGQIRASQSGSDVVIETDFGLRVAVDLVY 3774  
QY 446 Y---SV-----A---F---S-----PD----- 453  
Db 3775 YVRVTPGNYQLMCGLCGNYNGDPKDDFKPNSQAGNANEFQNSWEEVVPSPCLPPP 3834  
QY 454 -----G-----RY-----LAS-G---S-----FD--- 463  
Db 3835 TCPPGSEGCIPSEBCEPPELEKTKBEFCGLLSPTGPLSSCHKLVDPQGPLKDCIFDLC 3894  
QY 464 -----K-C-----V-----HI--W-N-T-----O-----T--G--AL 476  
Db 3895 LGGNLSILCSNIHAYVSAQAGGHVPEWRNETFCPMECPQNSHYELCADTCSLGSAL 3954  
QY 477 -----V-----HS--Y-----R----- 481  
Db 3955 SAPLQCPDCAEGCQCDGFLYNGACVPIQQCGYHNGVYVPEQTVLIDNCRQQTCH 4014  
QY 482 -----G-----TGIF-----E-----VC----- 490  
Db 4015 VGKVVQCQHSCKRPGQVCPSPGGILSCVKNKDPCHGVTCRPQETCKEQQGGVCLPNEYAT 4074  
QY 491 -----W-----N----- 492  
Db 4075 CWLWGDPHYHSFDRKFDFOGTCNYVLATTGCPGVSTQGLTPFTVTYKQNRGNPAVSYY 4134  
QY 493 -----AA-G-----D-----KV-----GAS-----A-----S 503  
Db 4135 RVVTVAALGTNISHTKDEICKVRVNGVLTALPVSVADGRISVAQASKALLVADFGLOVS 4194  
QY 504 -----D-----GSVC-----V-----L-----D-L 512  
Db 4195 YDMNRVDTVLPSSYHGAVCGLCGNMDRNPNDQVPFNGTLPSPITPIWGGSWRAPGDPL 4254

QY 513 -----R-----K 514

Db 4255 CWDECRGSCPTCE 4268

## RESULT 10

US-09-334-220-1  
; Sequence 1, Application US/09334220  
; Patent No. 6323177  
; GENERAL INFORMATION:  
; APPLICANT: St. Jude's Children's Research Hospital  
; APPLICANT: Curran, Thomas  
; APPLICANT: D'Arcangelo, Gabriella  
; TITLE OF INVENTION: INTERACTION OF REELIN WITH VERY LOW  
; TITLE OF INVENTION: DENSITY LIPOPROTEIN (VLDL) RECEPTOR FOR SCREENING AND  
; TITLE OF INVENTION: THERAPIES  
; FILE REFERENCE: 2427/0F704  
; CURRENT APPLICATION NUMBER: US/09/334,220  
; CURRENT FILING DATE: 1999-06-16  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 1  
; LENGTH: 3460  
; TYPE: PRT  
; ORGANISM: Homo sapien  
US-09-334-220-1

Query Match 69.7%; Score 2557.9; DB 3; Length 3460;  
Best Local Similarity 11.8%; Pred. No. 2e-32;  
Matches 401; Conservative 82; Mismatches 23; Indels 2901; Gaps 362;

QY 1 M-S-----I-----S-----S-----D-----EV----- 8  
Db 1 MERSGWAROTFLALLLGLATLRARAAGYYPFPFFLCTHHGELEGDEGEVLSLH 60  
QY 9 --N-----F---LV---YRY---LQ---ES-----GF---S-H- 23  
Db 61 IAGNPTYVPGQEHYHTISTSTFFDGLLVTLG-LYTSTVQASQSIGSSAFGFGINSDHQ 119  
QY 24 -----S-A-----F-----T-----F-----G--I-----E 31  
Db 120 FGNQFMCVAVSHVSHLPTTNLSFIWAPPAGTCGVNFMATATHRGQVIFKDALAQOLCE 179  
QY 32 -----S-----H-----I-----S---O---SNI-----N-----CAL----- 43  
Db 180 QGAPTDTVPHLAEIHSDSIILRDDFDYHQLQLPNIWVECNCTGEGCQGAIMHGNA 239  
QY 44 V---P--P---A-AL---I-----S-----SI-----IQ--- 54  
Db 240 VTFCEPYGPRELITTTGLNTTASVLQFSIGSGSCRFSDPSIIVLYAKNSADWQLEK 299  
QY 55 -----KG---LQY---V-BA-E-----VSIN----- 67  
Db 300 IRAPSNVSTIIHLYLPEDAKGENVQFQWKOENLRVGEVYEAACWALDNLII-INSARQV 358  
QY 68 --ED-----GT-LF-----DG-----R----- 76  
Db 359 VLEDSLDPVDTGNWLFPPGATVKHSCQSDNSIYFHGNESEFNFATRDVDLSTEDIQE 418  
QY 77 -----P-----IES-LSLI-----DA----- 86  
Db 419 QWSEEPESQPTGWDVLGAVIGTECGTIESGLSMVFLKDGKRLCTPSMDTTGYGNLRFYF 478  
QY 87 VM-----P-----D-----VV-----OT----- 94  
Db 479 VMGGICDPCGNSHENDIILYAKIEGRKEHITLDTLSYSSYKVPVSLVSWINPELOTPATKF 538  
QY 95 --RO---QAV-R---D---K-LAQ-----O----- 106  
Db 539 CLRQKHQGHNRNVAVDFHVLVPLPSTMSHMIQFSINLGCCTHGPNGSVSLEFSTNHG 598  
QY 107 -----HA-----A-----AAA-AA-----AAAT-N-----Q----- 120

Db	599	RSWSLLHTECLPEICAGPHLPHSTVYSSYENSGWNRITILPLNAALTRNRIRWQTGPI	658	Db	1678	NGSHVOLQYSLNNGKDWHLVTBECVPTTICGLHYTESSYTSRFRQWKRITVYVPLSTI	1737
Qy	121	-----G-S-----A-----K-----	125	Qy	252	-----W-----T-----	253
Db	659	LGNWAIIDNVYIGPCLKFCGRGQCTRHCKCDPGFSPACEMASQTFPMFISEFGSS	718	Db	1738	SPRTRPRWQANTVVGADSWAIDNVVLASGCPWMCGRGICDAGRCVCDRGFGPGYCPV	1797
Qy	126	-----N-----G-E-----N-----	129	Qy	254	-----KD-----GNL-----A-----S-T-L-----G-----	263
Db	719	RLSSYHNFYSIRGAESVFGCVLASGKALVFNKEGRRQLITSFLDSSQSRFLQTLRLGS	778	Db	1798	VPLPSILKDDFNGNLHPLDWPEVYGAERGNLGETIKSGTSLIFKGEGLRMLISRDLDCT	1857
Qy	130	-----T-A-----NG-----	133	Qy	264	-----Q-----HK-----G-----P-----I-F-----	270
Db	779	KSVLSTCRAPDQPEGVLLHYSYDNGITWKLLEHYSYLSYHEPRIISVELPGDAKQFGIQ	838	Db	1858	NTMYQFSRLRFIAKSTPERSHILLOFSISGGITWHLMDDEFYPPQTNNILFINVPLPYTA	1917
Qy	134	-----E-----E-----N-----G-----AH-----	139	Qy	271	-----A-----LKW-----N-KK-----GN-----F-----	280
Db	839	FRWQPHSSQREDVWAIDEIIMTWSLVFNSISLDFNLVEVTOSLGFYLGNVQPYCGHDW	898	Db	1918	QTNATRFRL-WOPYNGKKEBIWVDDFIIDGNVNNVNNVLLDTPDFGPREDNWFYPPGG	1976
Qy	140	-----T-----I-A-----N-----NH-----	145	Qy	281	-----I-L-----SA-----G-V-----D-----KTTII-----	292
Db	899	TLCFTGDSKLASSMRVYETOSMOIGASYMIQFSLVMCCGQKYTPHMDNQVKLEYSTNHGL	958	Db	1977	NIGLYCPYSSKGAPBEDSAMVFSVNEVGEHSITTRDLNVNENTIIQFEINVGCSTDSSEA	2036
Qy	146	-----D-----M-----M-E-----V-D-----G-----	153	Qy	293	-----W-----DAH-----T-----GEAKQ-----Q-----P-----	304
Db	959	TWHLVQEBCLPMPSCQEFTSASIHASEFTQWRVIVLLPQKTWSSATFRMSQSYTA	1018	Db	2037	DPVRLFEFRDGCATWHLHLLPLCYHSSSHVSSSLCSTEHPFSYIAG-TMQGMRREVHFG	2095
Qy	147	-----D-----M-----M-E-----V-D-----G-----	153	Qy	305	PEH-----SA-----PALD-----VDM-----Q-----S-----N-----T-----PA-----	323
Db	1019	QDEWALDSIYGOCNMCSGHGSCDHGICRCDOGYQGTCHPEAALPSTIMDFENQNG	1078	Db	2096	KUHLCSGVRFRWYQGYFAGSQPVTAIDNVYIGPOCEMNCQGSGCINGTKICIDPGYS	2155
Qy	154	-----D-----V-----E-I-----P-----S-----NKA-----V-V-L-----	165	Qy	324	-----SC-----ST-----D-----M-----C-----I-----	331
Db	1079	WESDWQEVIGGEIVKEPQCGVSSGLYFSGAKRQLVSWDLDTSWDFVQFYIQIGG	1138	Db	2156	GPTCKISTKNPDPLDKDFEGQLESDFLLMSGCKPSKCKGILSSGNNLPENEDGLRLMT	2215
Qy	166	-----R-----G-----H-----E-----SEV-----FI-----C-----	175	Qy	332	-----H-----V-----C-K-----L-G-----QD-----R-----	340
Db	1139	ESASCNKPSRERGVLQYSSNGGIQWHLAEMFYDFSKPRFVYLELPAAATPCTPR	1198	Db	2216	RDLDLSHARFVQFFMRLGCGKGVDPDRSQPVLLQYSLNGLSLQEFLLFSNSSNVGRY	2275
Qy	176	-----AW-----NPV-----S-D-----L-L-----	184	Qy	341	-----PIK-----T-----EQ-----GH-----T-----NEV-----NAI-----K-----	356
Db	1199	WQPVFSGEDYQWAVDDIILSEKQIIPVINTPLPQNFYKPAFYPMNQMVSVMML	1258	Db	2276	IALEIPLKARSSTRLRWQPSENGHYFSPWIDQILIGN-ISGNTVLEDDFTTIDSRK	2334
Qy	185	-----V-----S-----G-S-GD-----S-TAR-----I-----	195	Qy	357	W-----D-----P-----TG-----N-----LL-----ASCS-----D-----	369
Db	1259	ANEGMVNETFCAATSMIFGSDGDRFAVT-RDLTKPGYVLPKLNIGCANQFSSTA	1317	Db	2335	WLLHPGGTKMPVCGSDALVFIKASTRYVSTOVAVNEDSFLOIDFAASCSTVDSCTA	2394
Qy	196	-----WNL-----S-----E-NS-----T-----SG-----P-----	206	Qy	370	-----DM-----T-----L-KI-----W-----S-----	377
Db	1318	PVLLQYSHDAGMSWFLVKEGYPASAGKCGEGNSRELSEPTMYHTGDPBEWTRITIVIPR	1377	Db	2395	IBLEYSVDLGLSWHPLVRDCLPTNVECSRHYLQRIILVSDTFNKWTRITLPLPYTRSOAT	2454
Qy	207	-----T-----Q-----L-VL-----R-----H-----CI-----	215	Qy	378	-----M-----KQ-----DN-----CV-----H-----D-----	386
Db	1378	SLASSKTRFRWIOESSQKNVPPFGLDGVYISBPFCPSYCSGHGDCISGVCFDLGYTAAQ	1437	Db	2455	RFWRHQPAPDFKQQTWAIIDNVYIGDGCIDMCSGHGRCIOGNCVCDQMGGLYCDDPETS	2514
Qy	216	-----R-EG-----G-Q-----D-----D-----VP-----	223	Qy	387	-----L-----Q-----Q-----A-H-----NKE-----	393
Db	1438	GTCVSNVPHNEMFDFEGKLSPLWKITGAQVGTGCTLNDGKSLYFNGPGKREARTVP	1497	Db	2515	PTQLKDNFRAPSSQNWLTNGGKLTVCVAVASGAWALHFGSGCSRLLVTDNLNNAEF	2574
Qy	224	-----SN-----K-D-----	227	Qy	394	I-----Y-----T-----IKW-----S-PTG-----	403
Db	1498	LDRNIRLVQFYIQGSKTSGITCIPRTRNEGLIVQYSSNDNGILMHLLELDFMSFLEP	1557	Db	2575	IQFYFMYGCLITPNRNRNQVLLEYSVNGGITWNLMEIFYDQYQSKP-GFVNILLPPDAKE	2633
Qy	228	-----VTS-----L-D-----W-----N-----SE-----	236	Qy	404	-----P-----G-TN-----N-----P-NA-----	411
Db	1558	QIISIDLPODAKTATPAFRWQPHGKSAQWALDVLIGMNDSSQTGFQDKFGSIDLQ	1617	Db	2634	IATRFWRWQPRHDLQDNDWAIDNVLISGSADQRTVMTLDTFSSAPVPOHRSADAGPVG	2693
Qy	237	-----G-----T-L-----LA-----T-----G-----	243	Qy	412	-----N-----L-----ML-----A-----	416
Db	1618	ANWYRTOGGQVDICLSMDTALIFTENIGKPRYAEWDFHVSASTFLOEMSGCKSPFS	1677	Db	2694	RIAFDMFMDKTSVNEHWFHDDCTVERFCDSQDGVMLCGSHDGRVYAVTHDLTPEGW	2753
Qy	244	-----S-----Y-----D-----G-----FA-----RI-----	251	Qy	417	-----S-----A-----S-----F-----D-----S-----	422
				Db	2754	IMQFKISVGCKVSEKIAQNIHVQYSTDFGWSNVLVPOCLPADPKCSGVSQPSVFPT	2813

```

QY 423 ---T---VR---L---W-D---VD---RG---IC 434
Db 2814 KGWKRITVPLPESLVGNPFRFYQKYSMDQWAINDFYLPQCLNCRGHDCLREQCIC 2873
QY 435 ---I---HTL-T---KHQ---E-P---V---Y---S- 447
Db 2874 DPGYSGPNCYLTHLTKLKERFDEEIKPDLWMSLEGSTCTEGGILAEATLYFGST 2933
QY 448 ---V---A-F---S-P---DG---454
Db 2934 VRQAVTDLDLRGAKFLQWGRIGSNNMTSCHRPKRKEGVLLDYSDGIGITWLLHEM 2993
QY 455 ---R-Y-L---A---SG---460
Db 2994 DYQKISVRHDYILLPEDALNTTTLRWQPFVISNGIVSGVERAQWALDNILLIGGAEI 3053
QY 461 ---S---FD---K---C---VHI-W-N---T---471
Db 3054 NPSQLVDFDDEGTSHEENWSPFYNVATAGFCGNPSPHLYWPNKKDKTHNALSRELI 3113
QY 472 ---Q---T-G---A---IV---477
Db 3114 IQGYMMQFKIIVGCEATSCGDLHVMLEYTKDARSDSWQLVOTQCLPSSNSIGCSPPQ 3173
QY 478 -H--S-Y---R---G-T---483
Db 3174 FHEATIYNSVSSWKRTIOLPDHVSSATQFRWIOKBETEKSQWAIHDVYIGEACPK 3233
QY 484 ---G---G---I---FE-VC---W-N---492
Db 3234 LCSGHGYCTTGAICICDESFGDDCSVPSHDLPSYIKDNFESARVTEANWETIQGGVIGS 3293
QY 493 ---A-A-GKV---G---ASAS---D---GS-V-C---508
Db 3294 GCGOLAPYAHGDSLYFNGCQIRQA-ATKPLDLTRASKIMFVLQIGSMSTDCNSDLSP 3352
QY 509 ---VL---D---L-R-K 514
Db 3353 HAVDKAVILQYVNVNGITWHVIAHQPKDFQAOQSVNVNPLEARMK 3399

```

## RESULT 11

```

US-08-471-119A-2
; Sequence 2, Application US/08471119A
; Patent No. 5827706

```

## GENERAL INFORMATION:

```

; APPLICANT: Leitner, Ernst
; APPLICANT: Schneider, Elisabeth
; APPLICANT: Schoergendorfer, Kurt
; APPLICANT: Weber, Gerhard
; TITLE OF INVENTION: Cyclosporin Synthetase
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESS: No. 5827706artis Corporation
; STREET: 59 Route 10
; CITY: East Hanover
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07936

```

## COMPUTER READABLE FORM:

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,119A
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kassenoff, Melvyn
; REGISTRATION NUMBER: 26,389
; REFERENCE/DOCKET NUMBER: 100-8029/CONT/CONT

```

## TELECOMMUNICATION INFORMATION:

```

; TELEPHONE: 201 503 8474
; TELEFAX: 201 503 8807
; INFORMATION FOR SEQ ID NO: 2:

```

## SEQUENCE CHARACTERISTICS:

```

; LENGTH: 15281 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:

```

```

; ORGANISM: Tolypocladium niveum
; STRAIN: ATCC 34921
; US-08-471-119A-2

```

```

Query Match 69.5%; Score 2550.8; DB 2; Length 15281;
Best Local Similarity 7.6%; pred. No. 8.2e-29;
Matches 469; Conservative 39; Mismatches 4; Indels 5662; Gaps 376;

```

```

QY 1 M-----S-----ISSDE-----VNF-----L-----V 12
Db 5261 MKATNELSSRYAAVLHI-SDEPLPIYKIDPEAWINFEGRSLTREALAQVLKENENAESV 5319
QY 13 -----Y-----R-----YL-----Q-----17
Db 5320 AISNIPYKTVVERHIVRSILDQEDANAPESMDGSDWISAVRTRAQQCHTSLASDLFDIA 5379
QY 18 E-SGF---S---HS---A-F---T---F-----28
Db 5380 EDAGFVRSVSWARQHSQHGALDAVFHHLKPAEDSRVLKFPYTDHGRPLKSLTNQPLP 5439
QY 29 -----S-----G-----I-----E 31
Db 5440 AQSRRAELLIREGLQTLPPYMPISQITLIDRMLNANGKVDRLARRAKITQSKPVE 5499
QY 32 -----S-----H-----IS---35
Db 5500 DIVPPRNSVEATVCKGFTDVLGVGEVITDNFFNLGHSIMATKLAARLGRQLNTRISVRD 5559
QY 36 -----QSN---I-----N-GA-----42
Db 5560 VPDQPVVADLAAVIQNSAPHEPIKADYTPGVPSQFAQRLWFLDQLVGATWYLMPLG 5619
QY 43 -----L-----V---P-----P---46
Db 5620 IRLHSLRVDALATAISALEQRHEPLRTTFHEEDGVGVQVQDHRPKDLRIIDLSTQPKD 5679
QY 47 -----A-----ALI-----S---II-----53
Db 5680 AYLAVLKHEQTTLFDLATEPGHVRVALIRLGEHEHILSIWMHHIISDGWSVEVLFDHMRH 5739
QY 54 -----Q-K---GL-QY-----V-----E-----AEV---64
Db 5740 YSSALRQQDPMEQILPLPIQYRDFAAWQKTEQVAHQDLYWTEHLADSPAEILLTDL 5799
QY 65 ---SI---N-----ED---G 70
Db 5800 PRPSILSGRANBELPLTIEGRHLDKLRAFRCRVHQATPFVILALAAARAAHYRLTGAEDATLG 5859
QY 71 T-----L-----F---D-----74
Db 5860 TPIANRPRELENMIGFFVNTQCMRIAEENDNFSLVRVSTATSAFANQDVPFESIV 5919
QY 75 ---G---R-P---I-----E-----S-----80
Db 5920 SSLLPGSRDASRNPLVQVILAVHSQODLGLKLFLEGRDEAVDSAISTRFDVEFHLFEHAD 5979
QY 81 -LS-----L-I-----D-----AVMP-----DV 91
Db 5980 RLUSGVLYAKELFKLRTIESVVSFLETIRRALDQPLPLAVILPLTDTGCVGETASKGLLDV 6039

```

QY 92 -----V-Q-TR-----QQ----- 97  
Db 6040 PRDYPRDANIVEFQHQVREATDAJAVKDATSILTYAQLDQSDRLAIWLSRRHMPET 6099  
QY 98 -----AY-----R-----D-KL----- 103  
Db 6100 LVGLAPRSCETIIMFGIMKANLAYPLDINSAPAARLSILSAVDGNKLVLLGSGVTAP 6159  
QY 104 -----A-Q-----Q-----H----- 107  
Db 6160 EQENPEVEAVGIEIAGTGLDQKTSQSNAPPSATSILAYIFTSGSTGKPKGVMEHRSVT 6219  
QY 108 -----A-AA-A-A-AA-----AAT-N----- 119  
Db 6220 RLAKPSNVISKLPQGARVARHLANIAFDASIEWEATILLNGATVLCIDYHTVLDRCRTKEV 6279  
QY 120 -----Q-----Q----- 121  
Db 6280 FERESITVVTLPALKQCVAEIPETLAHLDDLTYTGGDRVGHDAAMRARSCLKIGMSGY 6339  
QY 122 G-----S-----A-----K-----N-----GE----- 128  
Db 6340 GPTENTVITIIYEDADEMFVNGPIGKTVNSGAYVMORNQOLVPSGVVGVGLVVTGDGL 6399  
QY 129 -----N-----TANGE----- 134  
Db 6400 ARGYTDPSLNKRFIYITVNGESIRAYRTGDRVYRPHDLQIEFFGRMDQOVKIRGHRIE 6459  
QY 135 -----E-----N-----GA----- 138  
Db 6460 PGEVESALLSHNSVQDAAVVICAPADQSDGAEMVAFVAARNTEDETOBEAEVDQVQGW 6519  
QY 139 H-T-----I-----ANN-----H-----T-----D----- 147  
Db 6520 THPETAAYSEVKDIROSEVGNDFMGWTSYDGSSEIDKTMHEWLNMTMILDAAREPGHV 6579  
QY 148 -----M-M-----E-----V-----DG-----D----- 154  
Db 6580 LEIGTGTGMFNLAKCPGLQGVGFPEPSKSAQFVNDAAQSPALKDGRSIVHVGTATD 6639  
QY 155 -----VE-----IPS-----N----- 160  
Db 6640 INKAGPIQRLVVINSVAQYFPTPEYLFRVVEALVQIPSVIRIVFGDMRTNAINRDFVAS 6699  
QY 161 -----KA-----V-V-----LR-G-G-----H----- 168  
Db 6700 RALHTLGEKANKRLVRQMIYELEANEHEELTDPAPFTSLRUGEKIKHVEILPKTMKAT 6759  
QY 169 -E-----S-E-----V-----F-----I----- 174  
Db 6760 NELSRYAAVLHVRSREQSTIHQVSPNAWIDFAADGLDRQTLINLLKEHKDAGTVAIG 6819  
QY 175 -----W-----N-----P-----V-SD----- 182  
Db 6880 GYQVEVSWARQWSONGALDAIFHHFPPKEGARTLIEFPTDYEGRNVTLTNRPLNSIOS 6939  
QY 183 -----LL-V-S-----G-----S----- 188  
Db 6940 RRLGTQIREKQLTLLPYMIPSRIMVLDQMPVNNNGKIDRKELVRAIVAPKPSAATRV 6999  
QY 189 -----G-----D-----S-----T-----AR----- 194  
Db 7000 APRNEIEAILRDEFVIGTEVSVLDNFFDLGCHSLMATKLAARVSRRLDAHISIKDVED 7059  
QY 195 -----I-----W-----N----- 198  
Db 7060 QPVLADLAASIORESAPHEIPORPYTGPAEQSFAQGRWFLDQLNLGATWYLMPLAIRI 7119  
QY 199 -----S-----E-N-----ST-SG----- 205

Db 7120 RGQLRVAALSAALFALERRHETLRTTFEESDGVGVQIVGEARNSDLRVHDSVGGDGEVL 7179  
QY 206 -----P-----T----- 207  
Db 7180 EVLRREQTVFFDLSEPGHVRCLVKTGEEDHVLISVMHHIIYDGSVDILRGELGQFYS 7239  
QY 208 -----Q-----L-V-LR-----H----- 213  
Db 7240 ALRGQDPLLANPLPIQYEDFAAQWQREAKQVEEHQQLGWSKQLVDSTPAELLTDLPRP 7299  
QY 214 -----C----- 215  
Db 7300 SILSGRAGSVDTIEGVSYGALQSCFRTSRVTTFFVLLTVFRIAHRFLTAVDATIGTPI 7359  
QY 216 -----R-E-G-----G-----QDVP----- 223  
Db 7360 ANNRPELTLVGCFFVNTQCMRISIAADDNFELVRQVRNVATAAYANQDVPFERIVSAL 7419  
QY 224 -----S-N-----K----- 226  
Db 7420 VPGSRNTRSNPLVQLMFAVQSVEDYDQVREGLESVMMPGEASTRFDMEFHLVPGQKLT 7479  
QY 227 -----D-----V-----TS 230  
Db 7480 GSVLYSSDLFEQGTIQNFVDIFQECILRSVLDQPLTPIISVLPFSNAISNLESLLDLEMP 7539  
QY 231 -----LD-----W----- 233  
Db 7540 DYPRDRTVVDLFREQAACPDIAVKDSSQLTYAQLDEQSDRVAALHERHMPAESLVG 7599  
QY 234 -----N-----S-EG----- 237  
Db 7600 VLSRSCETIIAYFGIMKANLAYPLDLYVAPDARLAAILDVTEGERLLLLGAGVPOPGIQ 7659  
QY 238 -----T-----L-----LA-----TG----- 243  
Db 7660 IPRLSYIAEALSHATTVDVTSIPOPATSILAYVIFTSGTGKPGVMIEHRGIVRLVR 7719  
QY 244 -----S-----Y----- 245  
Db 7720 DTNVNVFPGSGALPVSHFNSLAWDAATWEIYAVLNGTTCIDRDTWLDIAALNSTFR 7779  
QY 246 -----D----- 246  
Db 7780 KENVRAAFTPAFLKQCLAEETPELVANLEILHTAGDRDLPDGDANLAGKTAKGIFNVLGH 7839  
QY 247 -----G-F-----AR 250  
Db 7840 TENTAVSTFYPVVEETFFVNGVGRGISNSHAYIIDRHQKLVPAQVGMGELILTGQDVAR 7899  
QY 251 -----I-W-T-K-DG----- 256  
Db 7900 GYTDSALNKDRFVYIDINGKSTWSYRTGDKARYPRDQLEFFPRQMDQVKGIRGRIEPG 7959  
QY 257 -----LG-----Q-----HK----- 256  
Db 7960 EVELTLLDHKSVAATVVRRPFPNGDPEMIAFITDAEDDVQTHKAIYKHQILPAYMI 8019  
QY 257 -----N-LA-ST----- 261  
Db 8020 PSHLVLDQMPVTDNGKVDKDLALRAQTQVKRSTAAARVPPRDEAVEVLCEBYSNLEEV 8079  
QY 262 -----GP-----LKM-----N-----KKG----- 266  
Db 8080 EVGITDGFDLGCHSLATKLAARLSQNLTRVSVKDFDQPIADLADIIRRGSHRHP 8139  
QY 267 -----GPI-FA-LKM-----N-----KKG----- 279  
Db 8140 IPATPYTGPEQSFAQGR-L-WFLEQLNGASVWLMFPFAIRMRGPIQTKALAVNALVHR 8198  
QY 280 -----F-----I-LS-AGVD-----K-----TT----- 290

Db 8199 HEALRTTFEDHGVGVQVOPKSSQDLRIIDLSA-VDDTAYLAALKREQTAFDLTSEP 8257  
QY 291 -----II-----  
Db 8258 GWRVSLRLGDDDDYLSVMHHIISDGTVDVLRQELGQFYSAAIRGQEPLOSQAISLPQ 8317  
QY 293 -----W-----D-----294  
Db 8318 YRDPVWQROENIQEQAQKLYWSQQLADSTPCEFLTDLPSPILLSGEADAVPMVIDGT 8377  
QY 295 -----AH-TG-----E-----299  
Db 8378 VYQLLDFCRTHQVTSFVLLAAAFRTAHYRLTGTLDTAGTPIANRRNPELEGLIGFFVN 8437  
QY 300 -----A-K-QQ-----F-PF-----306  
Db 8438 TOCRMATSETFESLVQVRLTTTEAFANQDVFFEQIVSTLLPGSROTSRNPLVQVMF 8497  
QY 307 -----H-----S-----308  
Db 8498 ALQSQDLGRIQLEGMTDEALETPSLTRLDLVHLFOEVLKSGSLLYSTDLEFEVETIRG 8557  
QY 309 -----A-PAL-----D-----V-314  
Db 8558 IVDVLEILRRGLEQPKORLMAMPITDGTIKLRDQGLLTAKPAYPRESSVIDLFRQOVA 8617  
QY 315 -----D-W-----OSN-----N-T-----FA-SC-----325  
Db 8618 AAPDAIAVWSSSTLTVADLDGOSNLAHLWCORNMAPETLVAVFAPRSCLTIVAFGLVL 8677  
QY 326 -----325  
Db 8678 KANLAYPLDVPNAPARIEAILSAPVGHKLVIVQAHGPELGLTMADETVELOIDEALASS 8737  
QY 326 -----ST-----D-----M-329  
Db 8738 SGDHEQIHASGPTATSLAYVMFTSGSGKPKGVMIHRSIIRLVXNSDVVATLPTVRMA 8797  
QY 330 -----C-----I-----HV-C334  
Db 8798 NVSNLAFDISVQEIYTALLNGTIVCLDYLTLDISKILYNVVFVEAQVNAAMFTFVLKQC 8857  
QY 335 -----KL-----G-QD-----339  
Db 8858 LGNMPAIIISRLSVLFVNGDRDLDAHDAVASGLQDAVYVAYGTPENGQWSTMYKVDVNEP 8917  
QY 340 -----R-----P-----341  
Db 8918 FVNGVPIGRSITNSGAYVMDGNQOLVSPGVNGEIVVTGDLARGYTDSALDEDRFVHVTI 8977  
QY 342 -----IK-T-----P-O-----GH-----348  
Db 8978 DGEENIKAYRTGDRVRYPKDFIEFPGRMDQVKINGHRIEPAEVEHALLGHDLVHDA 9037  
QY 349 -----TN-----E-V-N-----A-IK-----356  
Db 9038 VVLRKANQEPMEAFITTSQDEETIEHESNKQVQWGEHFDVSRYADIKDLSTSTFGHD 9097  
QY 357 -----W-----D-P-----TG-NL-----363  
Db 9098 FLGWTSMYDGVIPVNMKEWLOETTASLLDNRPFGHILEIGAGTGMILSNLKGVDGLQK 9157  
QY 364 -----LA-----365  
Db 9158 YVGLDPAPSAAI FVNEAVKSLPSLAGKARVLVGTALDIGSLDKNEIQPELVINSVAQYF 9217  
QY 366 -----SC-----S-----D-----369  
Db 9218 PTSEYLLKVVAVVEPSVKRVFGDIRSQALNRDFLAARVALGDNASKEQIREKIAE 9277  
QY 370 -----D-----M-T-----L-----K-374  
Db 9278 LEESEELLVDPAFFVLSRSLPNKIKHIEVLPKLMKATNELSSRYAAVLHISHNEEQOL 9337

QY 375 -----I-W-----S-----M-----K-----QD-381  
Db 9338 LIQIDPTAWDFAATQKDSQGLRNLLQQGRDDVMIAVGNIPYSKTIIVERHIMNSLDQDH 9397  
QY 382 -----381  
Db 9398 VNSLDGTSWISDARSAAAICTSPDAPALTQAKEGFRVELSWARQSRONGALDAVFHRL 9457  
QY 382 -----NC-VH-D-----LQ-----388  
Db 9458 ATDANCERSRVLVHFPTDQGRQLRTLNRPLQASRRRIESQVFEALQATPALPAYMIPSR 9517  
QY 389 -----388  
Db 9518 IIVLPOMPTNANGKVDKQLARRAQVVAKRKAVSARVARNDETEIVLCBEYADILGTEVG 9577  
QY 389 -----388  
Db 9578 ITDNPFDMGHSILMATKLAARLSRRLDTRTVKEVDFKPVLDLAASIEQGSTPHLPAS 9637  
QY 389 -----A-----H-----390  
Db 9638 SVYSGPEQSYAQGRLWFLDQFNLNATWYHMSLAMRLLGLPLNMDALDVALRALEORHETL 9697  
QY 391 -----N-KE-----393  
Db 9698 RTTFEAKDQIGVOVHHEAGMKRLKVLDSKNEKEHMAVLENEQMRPFTLASPEGMKGHL 9757  
QY 394 -----I-YT-----I-----397  
Db 9758 ARLGPTETIILSLVHHMFSDGWSVDILRQELGQFYSAALRGDRDPLSQVKPLPIQYRDFAA 9817  
QY 398 -----K-W-----S-----P-T-GP-----404  
Db 9818 WQKEAAQVAEHERQLAYWENQOLADSTPGELLTDPPRPQFLSGKAGVIVPTIEGVPVYEKL 9877  
QY 405 -----N-N-P-----GT-N-N-P-----N-----410  
Db 9878 KFSKEROVTLFSVLLTAFRATHRPLTGAEDATIGTPIANRPELHIIIGFFVNTQCMRL 9937  
QY 411 -----A-N-----LM-----414  
Db 9938 LLDTGSTFESLVQHVRSVATDAYSNQDIPPERIVSALLPGSRDASRPLIQLMFALHSOP 9997  
QY 415 -----L-A-----S-A-----S-F-420  
Db 9998 DLGNITLEGLEHERLPTSVATRPFDMEFHLFQEPNKLSGSILFADELFPETINSVTVFQ 10057  
QY 421 -----D-ST-----V-R-----425  
Db 10058 EILRRGLDQPOVSIISTMPLTDGLIDLEKLGLEIENSPRDSYVVDVFRQOVAANPNAP 10117  
QY 426 -----L-W-----427  
Db 10118 AVVDSETSMYTSLDQKSEQIAAHLHAQGLRPESLIVMAPRSFETIVSLFGILKAGAY 10177  
QY 428 -----DV-----DR-----431  
Db 10178 LPLDVNSPAARIQIPILSEVEGKRLVLLGSGIDMPOQSDRMDVETARIQDILTNTKVERSDP 10237  
QY 432 -----431  
Db 10238 MSRPSTSLAYVIFTSGSTRPKGVMIHNRNLRVVKQSNVTSQLPQDLRMAHISNLAFD 10297  
QY 432 -----G-ICI-----H-----436  
Db 10298 ASIWEIFTAILNGGALICIDYFTLLDSQALRTTFEKARVNATLFAFALLKECLNAPTFL 10357  
QY 437 -----436  
Db 10358 EDLKVLYIGDRLDATDAAKIQALVKGTVYVAYGPTENTVMSTIYRLTGESYANGVPIG 10417



QY 437 -----TL----- 438  
Db 10418 NAVSSGAYIMDKORLPPVGMGELVSGDLGARGVTNSTLNADRFVDIVINDOKARAY 10477  
QY 439 -----T-----K--HQ--FP----- 444  
Db 10478 RTGDRTRYRKPQGSIEFFGRMDQOVKIRGRHVEPAEVEQAMLGNAKHAIDAAVVQVAVDQG 10537  
QY 445 -----V-----Y-----S-- 447  
Db 10538 ETEMIGFVSNASDRFSEGEBEETNOQWEDHFEFESTAYAGIEAIDQATLGRDFTSWTSMY 10597  
QY 448 -----V-----A-----F-----S--P-- 452  
Db 10598 NGNLIDKAEMEELDDTMQSLDKEDARPACAEIGTGTGMVFNLPKNDGLGESVYGIPEPSR 10657  
QY 453 -----D-----G-----R-YL--A 458  
Db 10658 SAALFVDKAAQDFGLQGTQIILVGTABDIKLVKDFHPDVVVINSVAQVPPSRYSYLQVIA 10717  
QY 459 S-----G--S-----F-----DK----- 464  
Db 10718 SELHMTSVKTIFFGDMRSWATNRDFLVSALYTLGDKATKQIQEVARLENEDELLV 10777  
QY 465 -----W-----NT-----C-----VH-I 468  
Db 10778 DPAPFTSLTSQWPKGVKVEILPKMRMTSNELSSRYAAVLHICRDGEGNRYGRRVHSV 10837  
QY 469 -----W-----T-----S-----Y----- 471  
Db 10838 EENAWIDFASGMDRHALVQMLDERRDAKTVAIGNIPHSNTINERHFTTSLDTEGEGIAQ 10897  
QY 472 -----Q-----T-----GAL--V--H-- 478  
Db 10898 DSLDGSAMQSAKAMARCPCLSVTELVEIGQAAGFRVSVWARQSRQSGHALDVVFHLE 10957  
QY 479 -----S-----Y----- 480  
Db 10958 DDRVGRVLNFPDTERLPBPGTGLSRPLQRIQNRREFSQIREQLQTLPLPYVWPVSRIVV 11017  
QY 481 -----R-----G-T----- 483  
Db 11018 LERMPNANSKVDKELARKARTLOTIKPSATRVAPRNDIEAVLCDFQAVLGVTVGVMD 11077  
QY 484 -----G-----R-----B----- 488  
Db 11078 NPFELGSHLMATKLAARLSRRLDTRVSVKDIENQPIQLADVVTGSGAPHEAIFSTPY 11137  
QY 489 -----VC-----W-----NA-----A-----G----- 495  
Db 11138 SGPEQSQSGRLWFLDQLNLNASWYHMPPLASRLRGLRTEALQSLATIEARHESLRTT 11197  
QY 496 -----D-----K-----V-G-----A--SA----- 502  
Db 11198 FEBQDGVVPQIVRAARKQLRIIDVSGTEDAYLAALKQEODAFDLTAEPGMRVALLRLG 11257  
QY 503 -----SDG--SV----- 507  
Db 11258 PDDHVLIVMWHIISDGSVDILRQELGQLYSNASSQAPLPIQYRDDPAIWQKDSQIAE 11317  
QY 508 -----L-----D-----LR--K 514  
Db 11318 HQQLNWKRLQVNSKPAELLADPTRPKALSGDADVIPIBIDQVYQNLRSFCRARHVT 11377  
QY 509 -----V----- 11430  
Db 11378 FVALLAAFRAAHYRLTGAEDATIGSPIANRRRPELEGICGFVNTQCLRIPEK 11430

RESULT 12

US-09-052-469-8  
; Sequence 8, Application US/09052469  
; Patent No. 6380360

GENERAL INFORMATION:  
APPLICANT: Harris et al.  
TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE  
TITLE OF INVENTION: AND USES THEREOF  
NUMBER OF SEQUENCES: 23  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff, Ltd.  
STREET: One Financial Center  
CITY: Boston  
STATE: MA  
COUNTRY: US  
ZIP: 02111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk, 3.50 inch  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: WordPerfect 6.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/052,469  
FILING DATE: Concurrently herewith  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/422,582  
FILING DATE: 14-April-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9507766.5  
FILING DATE: 13-APR-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9411900.5  
FILING DATE: 14-JUN-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB PCT/GB94/02822  
FILING DATE: 23-DEC-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9326470.3  
FILING DATE: 24-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Williams, Ph.D., Kathleen M.  
REGISTRATION NUMBER: 34,380  
REFERENCE/DOCKET NUMBER: 3265/74165  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 345-9100  
TELEFAX: (617) 345-9111  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4302 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-09-052-469-8

Query Match 69.5%; Score 2548.4; DB 3; Length 4302;  
Best Local Similarity 9.8%; Pred. No. 1e-31;  
Matches 412; Conservative 83; Mismatches 13; Indels 3676; Gaps 357;

QY 1 MSIS-----S-----D-----EVN-----F----- 10  
Db 72 LDVSHNLLRALDVGLLANLSALAELDISNNKISTLEEGIFANLNLSEINLSPGECDC 131  
QY 11 -LVY--RY-----L--Q-----E-----SG-- 20  
Db 132 GLAWLPRWAEBOQVRYVQPEAATCAGPGSLAGPLLGIPLLDSCGCEYVACLPDNSSGT 191  
QY 21 ----PS--H-----SAF----- 26  
Db 192 VAAVPSAAHEGLLOPEACSAFCFTGCGLAALSEOGWCLCGAAQPSASFACLSLCSGP 251  
QY 27 -----T-----F-----G-----I----- 30  
Db 252 PPPAPTCTGPTLLQHVFPASPATGATLVGPHGPLASGLAAFAHIAAPLVATATRWDFDGS 311  
QY 31 -E-----SH-----I-----S-QSN-- 38  
Db 312 AEVDAAGPAASHRYVLPRGYHVTAVLALGAGSALLGTDVQVEAAPAALELVCPSSVQSD 371

QY 39 -I-N-G-AL-V-P-P-46  
Db 372 SLDSIQNRGGGLEAAYSIVALGEEPARAVHPLCPSDTEIFPONGHCYRLVVEKAWLQ 431  
QY 47 -AA-LIS-I-IQ-K-G-L-57  
Db 432 AQSCQAWAGALAMVDSPAVQFLVSRVTRSLDMWIGFSTVQGVGEPAPQGEAFSLES 491  
QY 58 Q-IV-E-AE-V-S-65  
Db 492 CONWLPGEHPATAEHCVRLGPTGMCNTDLCSAPHYSVCELPQGGPVQDAENLLVAPSG 551  
QY 66 -IN-EDG-TL-F-73  
Db 552 DLOGLPTPLAQDGLSAPHPVEVMPFGLRLGREAPLTTAEFGTOELRRPAQLRLQVVR 611  
QY 74 -D-GR-P-I-78  
Db 612 LLSTAGTPENGSEPSRSPDNRTQLAPACMPGGRWCEGANICLPDASCHPOACANGCTS 671  
QY 79 -E-S-LSLI-DAV-M-P-DVV-92  
Db 672 GPGLPGAPYALWREFLFSVAAGPPAQS-VTLHGQD-VLMLPGDLVGLQHDAGPGALLHC 729  
QY 93 -Q-T-R-Q-96  
Db 730 SPAPGHPGPOAPYLSANASSWNLPHLPAQLEGTWACAPACALLRLAATEQLTVLLGLRPNPG 789  
QY 97 -QA-Y-RD-KL-A-Q-105  
Db 790 LRMPGRYVRAEVGVSRRHNLSCSPDVSVPAGLVIYPAPRDGRLYVPTNGSALVLQV 849  
QY 106 -OH-107  
Db 850 DSGANATATARPWGVSARFENVCPALVATFVPGCPWETNDTLFSVALPMLSEGEHV 909  
QY 108 -A-A-A-A-A-A-113  
Db 910 DVVENSASRANLSLRVTAEEPIGRLATPSPEARVLQGVLYRSPVVEAGSMVFRWTI 969  
QY 114 -AA-A-AT-N-Q-Q-121  
Db 970 NDQSLTFQNVFNVIQSAAPKLSLTASNHSVNTVNTVVERNMRQGLQVSTVPA 1029  
QY 122 -G-SA-K-N-GE-N-TA-131  
Db 1030 VLSFNATLVLTGVLVDSAVEAFLWNFGDGOALHQFQPPYNESFPVDPDSVAQVLVEH 1089  
QY 132 N-GE-EN-N-G-A-H-139  
Db 1090 NVMTYAAPGEYLLTVLASNAFENLTQOVPSVRASLPSVANVGSDGVLVAGRPTFYPH 1149  
QY 140 -T-I-ANNHT-146  
Db 1150 PLSPGGLVLTWDFGDSPLVTSQPAANHTYASRGTYHVRLEVNNTVSGAAAQADRVF 1209  
QY 147 -DM-M-E-V-D-GD-154  
Db 1210 BELRLSLVDMSLAVEQCAPVVSAAVQTGDNITWTDFMDGDTVLSGPEATVEHVYLRQ 1269  
QY 155 -VS-IPS-162  
Db 1270 CTVTVGAASPAHLSRLHLVFLVLRVEPAACIPTQDARLTAYVTCNPAHYLFDWT 1329  
QY 163 -V-L-R-G-H-ESEV-172  
Db 1330 FGDSSNTTVRGCTVTHNTRSGTFLALVLSRVNRHAFYSICEPEVGNVTLQPER 1389  
QY 173 -FI-CAW-NP-179  
Db 1390 QFVQLGDEALVACAWPPFPYRYTWDFTGTEAAPTARGPEVTFIYRDPGSLVTVTASN 1449

QY 180 -VS-D-L-LV-S-GS-190  
Db 1450 NISAANDSALVEQVPEVPLVTSIKVNGSLGLELOQPFLFSAVGRGRPASVYLDLGDGWL 1509  
QY 191 -S-TARI-WN-L-SE-NST-SG-206  
Db 1510 GPEVTHAYNSTGDFTVRVAGWNEVSRSSEAWLNVTVKRRVRGLVVNASRTVPLNGSVSFS 1569  
QY 207 T-QL-VL-RHC-I-RE-G-G-OD-221  
Db 1570 TSLEAGSDRYSNVLCDR-CTPIPGGTISYTFRSVGTFTNIIVTAENVEGSAQDSIFVYV 1628  
QY 222 -V-PSN-KD-V-T-SL-231  
Db 1629 LQLEGLQVVGGRYPFTNHTVQLOAVVRDGTNVSYSWTAWDRGPALAGSGKFSLTVL 1688  
QY 232 -D-W-N-S-235  
Db 1689 EAGTYHVQLRATNMLGSAWADCTMDFVEPVGLMVTASPNPAVNTSVTLASLAGSGV 1748  
QY 236 -EG-T-L-LAT-GS-YD-246  
Db 1749 VYTWSLEELGSWETSEPTTHSPTPLGLHLVTMTAGNPLGSANATVEVDVQVPVSGLSIR 1808  
QY 247 -G-F-A-R-I-W-TK-D-G-NL-258  
Db 1809 ASEPGGSFVAAGSSVPFWGLATGNTVSWCWA VPGSSKRGPHVTMVFPDAGTFSLRLN- 1867  
QY 259 AS-T-L-QO-H-K-G-267  
Db 1868 ASNAWSVSATYNTLTAEEPIVGLVWASSKVWAPQOLVHFQVLLAAGSAVTFLQVGGAN 1927  
QY 268 P-I-F-AL-K-W-N-KG-278  
Db 1928 PEVLPGFRESHSPRVGDHVSVRGKNVSWAQAVRIIVLEAVSGLQMPNCEPGIATG 1987  
QY 279 -NF-ILS-AG-285  
Db 1988 TERNFTARVORGSRAVAYNFSLQKQVQDSLVLSGRDVTYTPVAAGLLEIQVAFNALG 2047  
QY 286 -V-D-K-T-T-II-WD-294  
Db 2048 SENRTLVLVEQDAVQVALQSGPCFTNRSAQFEAATSPSRVAYHWDFGDGPQDQDE 2107  
QY 295 -A-HT-G-E-A-KO-Q-FP-305  
Db 2108 PRAEHSYLRPGDYRVQVNASNLVSFFVAQATVTVQVLACREPEVDVVLQVLMRRSQRN 2167  
QY 306 F-H-SA-P-AL-DV-314  
Db 2168 YLEAHVDLRDCVTYQTEYRWEVYRTASCORPGPARVALPGVDVSRPLVLPRLALPVGH 2227  
QY 315 -D-W-QS-N-319  
Db 2228 YCFVFWVSGDPLTQSIQANVTAPERLVPPIEGGSYRVWSDTRDLVDGSESDPNLE 2287  
QY 320 -N-TF-322  
Db 2288 DGDQTPLSFWACVASTOREAGCALNFGPRGSSTVTIPRERLAAGVEYTFSLTVWKAGR 2347  
QY 323 -A-SC-325  
Db 2348 KKEATNQTVLIRSGRVPVLSLVCVSCKAQAVYEVSRSSVYVLEGRCLNCSSGSKRGWAA 2407  
QY 326 -S-T-D-M-C-I-331  
Db 2408 RTFSNKTLLVLDETTSTGSGAGMRLVLRGVRDGEYTFTLVLRSGBEEGCASIRLSP 2467  
QY 332 -H-V-C-334  
Db 2468 NRPPGSGSRLPPLGAVHALTTKVHFECTGWHDAEDAGAPLVYALLRLRCRQGHCEFCV 2527  
QY 335 -K-L-G-QD-P-I-342

Db 2528 YKSLSSYGA VLPGRPHFVGLVAVVQDQGAVALNRSIAITLPEPNGSATGLTVM 2587  
QY 343 ---KT---FG---H---T-NE---V---N---353  
Db 2588 LHGLTASVLPGLLRQDPOHVIBYSALVTVLNEYERALDVAABPKHERQRAIRKNT 2647  
QY 354 ---LA---S---C---SD---369  
Db 2648 ETIVSLRVHTVDIOQIAAALACQMGPSRELVCRCCLKQTLHKLEAMMLILOAETAGTV 2707  
QY 357 ---WD---PT-G---NL---363  
Db 2708 TPTAIGDSILNITGDLHLASSDVRAPOPSSELCAESPSRMASQAYNLTSALMRILMRSR 2767  
QY 364 ---LA---S---C---SD---369  
Db 2768 VLNEEPLTLAGEIVAQGRSDPRSLCYGAPGCHFSIPEAFSGALANLSDVVQLIF 2827  
QY 370 ---D---M-T---L---KI---W---376  
Db 2828 LVDSNPFPGYISNYTVSTKVASMAFQTQGAQIPIERLASERAITVKVPNNSDWAARGH 2887  
QY 377 ---SM---KQ---D---381  
Db 2888 RSSANSANSVVQPQASVGAVVTLDSNPAAGLHLQNLVYTLIDGHYLSSEPEPYLAVYLH 2947  
QY 382 ---NC---V---H---DLOA---H-N---K---EI 394  
Db 2948 SEPRNEHCASRRIRPESLOGADHRPYTFFISGSRD-PAGSYHLNLSHFRWSALQV 3006  
QY 395 ---YT---IKW---SP---T-G---403  
Db 3007 SVGLYTLSCQYFSEEDMV-WRTGELLPLEBTSRQAVCLTRHLTAFGASLFPVPPSHRVFV 3065  
QY 404 ---P---GT---N---404  
Db 3066 FPEPTADVNIYMLTCAVCLVYVMAAILHLKLDQDASGRRAIPFCGQGRPKYELVYK 3125  
QY 405 ---N---P---N---AN-L-M-L-AS-AS---419  
Db 3186 HDNKGSLSPAFLQHVIVRDLQARSAPFLVNDWLSVETEANGVLVEKEVLAASDAALLRF 3245  
QY 420 ---PD---S-T-V-R---L---W---D-428  
Db 3246 RRLVLAELQGFDDKHILWLSINDRPPRSRTRIQATCCVLLICLFLGANAVWYGA VGS 3305  
QY 429 ---VD---R---G---432  
Db 3306 AYSTGHVSRSLPSLSDTVAVGLSVVVPVYLAILFLFRMSRSKVAGSPSPAGQVQL 3365  
QY 433 I-C-I---H---TLT---439  
Db 3366 DIDSCLDSSVLDSSFTFSGLHAEAFVGMKSDFLDSSKSLVCWPSGEGTSLWPDLSSD 3425  
QY 440 ---K---440  
Db 3426 PSIVGSNLRLQARGHGLGPEEDGFSLASPYSPAKSPASDEDLIQVLAEGVSSPAP 3485  
QY 441 ---H---Q---E---443  
Db 3486 TQDTHMETDLSLSTPGKTTETLALQRLGELGPPSPGLNWQPOQAARLSRTGLVEGLR 3545  
QY 444 ---P---V-Y---SVA---449  
Db 3546 KRLLPAWCASLAHGLSLLLVAVAVAGVWGVGASFPFGVSVAMLLSSSASFLASFLGWEP 3605  
QY 450 ---PS---PD---453

Db 3606 KYLLEALYFSLVAKRIHPDEDDTLVESPAVTPVSARVPRVRPHGFALFLAKEEARKVKR 3665  
QY 454 ---G-R---Y---LAS---G---S-F---462  
Db 3666 LHGMLRSLVYMLFLVLTLLASVGDASCHGHAYRLOSATKQELHSAFLAITRSEELWPW 3725  
QY 463 ---DK---C---V---466  
Db 3726 MAHVLLPYVHGNQSSPELGPRLRQVRLQALYDPDPPGPRVHTCSAAGGFSTSDYDVGWE 3785  
QY 467 ---H---I---W-N-470  
Db 3786 SPHNGSGTWAYSA PDLLGAWSGSCAVYDGGVQBELGSLSEBSRDLRFLQLHNLMDNR 3845  
QY 471 ---TQ---T---G---AL---476  
Db 3846 SRAVFLRLTRYSPAVGLHAAVTLRLEFPAGRALAALSVRPPFALRLSAGLSPLLT SVC 3905  
QY 477 ---VH---SY---R---G---T---483  
Db 3906 LLLFAVHFAVAEARTWHREGRWRLGAWARWLLVALTAATALVRLAQGAADROWTRF 3965  
QY 484 ---G---G---IP---E-488  
Db 3966 VRGRPRFTSFDQVAHVSSAARGLAASLLFLLLVKAQHVRFVROWSVFGKTLCRALPEL 4025  
QY 489 ---V---489  
Db 4026 LGVTGLVGLVWAYAQLAILLVSSCVDSLWSVAQALLVLCPTGLTGLTCLCPAESWHLSPLL 4085  
QY 490 C---WNA---A-G---D---496  
Db 4086 CVGLMALRLWALRGAVILRWYHALRGELYPAPEDYENVELFLRLRLMGLSKV 4145  
QY 497 ---KV---G---A-S-AS---DG-SV-507  
Db 4146 KEPRHKVRPEGMPLPSRSRSGSKVSPDVPFPSPAGSDASHPSSTSSQLDGLSVSLGR LGT 4205  
QY 508 -C---V---L---D---L-RK 514  
Db 4206 RCEPEPSRLQAVFEALLTOFDRLNQATDVOYLEOQLHSLQGRR 4249

RESULT 13  
US-08-422-582-8  
; Sequence 8, Application US/08422582  
; Patent No. 6485960  
; GENERAL INFORMATION:  
; APPLICANT: Harris et al.  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: 75 State Street  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk, 3.50 inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/422,582  
; FILING DATE: 14-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9411900.5  
; FILING DATE: 14-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB PCT/GB94/02822  
; FILING DATE: 23-DEC-1994

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9507766.5  
; FILING DATE: 13-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9326470.3  
; FILING DATE: 24-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams, Ph.D., Kathleen M.  
; REGISTRATION NUMBER: 34,380  
; REFERENCE/DOCKET NUMBER: 3265/53313 (MRC-006xx)  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 345-9100  
; TELEFAX: (617) 345-9111  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4302 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-422-582-8

Query Match 69.5%; Score 2548.4; DB 4; Length 4302;  
Best Local Similarity 9.8%; Pred. No. 1e-31;  
Matches 412; Conservative 83; Mismatches 13; Indels 3676; Gaps 357;

QY 1 MSIS-----S-----D-----EYN-----F-----10  
DB 72 LDVSHNLLRALDVGLLANLSALAEIDISNNKISTLBEIGIPANLFNLSINLSGNPFECDC 131  
QY 11 -LVY--RY-----L--Q-----E-----SG--20  
DB 132 GLAWLPRWAEEOQVRVQVPEAATCAGPGSLAGPLGILLDSCGEEYVACLDPNSSGT 191  
QY 21 ----FS--H-----SAF-----G-----I-----26  
DB 192 VAAVSFAAHEGLQPEACSAFCSTGQGLAALSEQGWLCLGAAQPSASFACLSLCSGP 251  
QY 27 ----T-----F-----G-----I-----30  
DB 252 PPPAPTCTRGPTLLQHVFPASPGATLVGPHGLASQOLAFAHIAFLPVTAIRWDFGDS 311  
QY 31 -E-----SH-----AL-----V-----P-----I--S-QSN--38  
DB 312 AEVDAAGPAASHRYVLGRVHTAVLALGAGSALLGTDVQVEAAPALELVCSFVSQDE 371  
QY 39 ----I--N--G-----AL-----V-----P-----46  
DB 372 SLDLISQNRGGSGLEAAYSIVALGEEPARAVHPLCPSTDEIFPGNGHCYRLVVEKAALQ 431  
QY 47 ----AA-----LIS-----I-----IQ--K-----G-----L--57  
DB 432 AQEQCAWAGALAMVDSPAVQRFVRSVTRSLDVMIGFSTVGQVEGPAQGEAFSL 491  
QY 58 -Q-----YV-E-----AE--V--S--65  
DB 492 CQNLWLPCEHPATAEHCVRGLGPTGNCNTDLCSPHYSVCELOPGGPVQDAENLLVGAPSG 551  
QY 66 ----IN-----EDG-----TL--F-----73  
DB 552 DLQGLTPLAQDGLSAPHEPVEVMPFGLRLSREAFLLTAETGQELRRPAQLRLQVVR 611  
QY 74 ----D-----D-----CR--P--I-----78  
DB 612 LLSTAGTPENGSEPSRSPDNRQTOLAPACMPGGRWCPCGANICLPDASCHPQACANGCTS 671  
QY 79 ----E--S-----LSLI-----DAV-M-P-DVV-----92  
DB 672 GPGLPGAPYALWREFLESVAAGPPAQS-VTLHGQD-VLMLPGDLLVGLQHDAGPGALLHC 729  
QY 93 ----Q-----T-----R-----Q-----96  
DB 730 SPAPGHGPOAPYLSANASSWLPPLPAQLLEGTWACPCACALRLILAATEQLTVLLGLRPNPG 789

QY 97 ----QA-----Y-----RD-KL-----A--Q--105  
DB 790 LRMPGRYEVRAEVGNGVSRHNLSCSFDVVS PVAGLRIYIPAPRDGRLYPTNGSALVLQV 849  
QY 106 ----A-----A-----A-----A-----OH--107  
DB 850 DSGANATATARPWGGSVSARFENVCPALVATFVPGCPMETNDTLFSPWALPMLSEGEHV 909  
QY 108 ----A-----A-----A-----A-----A-----113  
DB 910 DVVENSASRANLSLRVTAEEPIGRLATPSPPEARVLQGLVRYSPVWAGSDMVRFTI 969  
QY 114 ----A-----AA-----AT--N--Q--Q-----121  
DB 970 NDKQSLTFQNVVFNVIYQSAANVFKLSLTASNHSNVTVNVTVERMRNMQGLQVSTVPA 1029  
QY 122 ----G-----SA-K--N--GE-----N-----TA-----131  
DB 1030 VLSPNATLVLTGGVLVDSAVEVAFLWNFGDGEALHOFPPYNESFPVDPSPVAQVLVEH 1089  
QY 132 N-----GE-----EN-----G-----A-----H 139  
DB 1090 NVMHTYAAPGEYLLTVLASNAFENLTQQVPSVRASLPSVAVGSDGLVAGRPVTFYPH 1149  
QY 140 ----T-----I-----ANNHT-----146  
DB 1150 PIPSPGGVLYTWDFGDSPLVTSQSPAAHNTYASRGTYHVRLEVNNTVSGAAAQADRVF 1209  
QY 147 ----DM-M--E-----V-----D-GD-----154  
DB 1210 EELRGLSDVMSLAVEGAPVWVSAAVQTGDNITWTGMDGDTVLSGPEATVEHVYLAQN 1269  
QY 155 ----V-----VE--IPS-----NKA-----162  
DB 1270 CTVTGAASPAGHARLSHLVFLVLEVRVEPAACIPTQDARLTAYVTGNPAHYLFDWT 1329  
QY 163 ----V-----V--L-R-G-----H-----ESEV-----172  
DB 1330 FGDGSSNTVRCGPTVTHNFTSGTFPLALVLSRVNRAHYFTSICVEVEGVNVTLOPER 1389  
QY 173 -FI-----CAW-----NP-----179  
DB 1390 QFVQLGDEAWLVACAWPPFPYRYTWDFGTEAAPTARGPEVTFIYRDPGSLVTVTASN 1449  
QY 180 -VS--D--L-----LV-S--GS-----GD-----190  
DB 1450 NISAANDSALVEQEPVLVTSIKVNGSLGLELQOQPYLFSAGVGRPASVYLDLGDGWL 1509  
QY 191 ----S-----TARI--WN-L--SE--NST-----SG-----P-----206  
DB 1510 GPEVTHAYNSTGDFTVRVAGNNEVSRSEAWLVTVKVRGLVNVNASRTVVPPLNGSVSFS 1569  
QY 207 T-----OL--VL--RHC-----I-----RE-G-----G--QD-----221  
DB 1570 TSLEAGSDVRYSWLCDR-CTPIPGGPTISYTFRSVGTENIIIVTAENEVSGAODSIFVYV 1628  
QY 222 ----V-----PSN-----KD--V--T-----SL-----231  
DB 1629 LQLIEGLVVGGRVFPFTHVTQQLAVRDGNTVSVSWTAWDRGPPALAGSGKGFSLTVL 1688  
QY 232 ----D-----D-----W-----N-----S--235  
DB 1689 EAGTHVQLRATNMLGSANADCTMDPVEPVGMLVWTASPNPAAVNTSVLSAELAGSGV 1748  
QY 236 ----EG-----T--L--LAT-----GS-----YD-----246  
DB 1749 VYTWLEEGLSWETSBPFTTHSGFTPLGLHLVMTAGNPLGSANATVEVDVQVPVSGLSIR 1808  
QY 247 ----G-F-A-----R-----I--W-----TK--D-G-----NL 258  
DB 1809 ASEPGGSFVAAGSVFVFWGQLATGNTVSWCNAVPGSGSKRGPHVTWFFDAGTFSIRLN- 1867  
QY 259 AS-----T-----L-----GQ--H-----K-----G--267

Db 1868 ASNAVSVSATYNLTAEPIVGLVWASSKVAPGOLVHFQILLAAAGSAVTFRLQVGGAN 1927  
QY 268 P-I- -F--AL- -K--W- -N--KG- -278  
Db 1928 FEVLPGFRFHSPPRGDHHVSVRGKHHVSAQAQVRIVVLEAVGLQMPNCCPEGIATG 1987  
QY 279 --NF- -ILS- -AG- -285  
Db 1988 TERNFTARVQRGSRAVAYFSLQKVGDSLVLSGRDVTYTPVAAGLLEIQVRAFNALG 2047  
QY 286 --V-D- -K--T- -T- -II--WD- -294  
Db 2048 SENRTLVLEQDAVQVVALQSGCFNRSQAPEAATSPSPRVAYHWDGDSGPGQDTDE 2107  
QY 295 --A-HT- -G--E--A-- -KO--Q-- -FP- -305  
Db 2108 PRAEHSYLRPGDVRVQVNASLVSFVFAQATVTVQVLACREPEVDVVLQVLMMRRSQRN 2167  
QY 306 F--H- -SA- -P--AL--DV- -314  
Db 2168 YLEAHVLDRCVYQTEYRWEVYRTASCQPPGRPARVALPGVDVSRPRLVLPRLALPVGH 2227  
QY 315 -----D-----W-----QS--N-- 319  
Db 2228 YCFVSVSGDTPLTQSIQANVTVAPERLVPIIEGGSYRWSDTDLVLDGSESDPNLE 2287  
QY 320 -----N-----TF----- 322  
Db 2288 DGOPTLSPHMACVASTQREAGCALNFGPRGSSVTIIPRERLAAGVEYTFSLTVWKAGR 2347  
QY 323 --A-----SC----- 325  
Db 2348 KEBATQTVLIRGRVPIVLESCVSCAKQAQVYEVSRSSVYLEGRCINCSSGSKGRWAA 2407  
QY 326 --S--T--D-----M-----C--I----- 331  
Db 2408 RTFSNKTVLDETTTSGAGMELVLRGVRDGEVGTFTLVLRSGSEEGCASIRLSP 2467  
QY 332 -----H-----V-----C-- 334  
Db 2468 NRPLGSCRLPPLGAVHALTTKVHFECTGWHDAEDAGAPLVYALLRLRCRQGHCEFCV 2527  
QY 335 -K-----L-----G-----OD-----R-----P-----I-- 342  
Db 2528 YKGLSSYGAIVLPPGRPFHFVGLVAVVQDLGAAVVALNRSIAITLPEPNSATGLTVW 2587  
QY 343 --KT--FQG-----H-----T--NE-----V-----N----- 353  
Db 2588 LHGLTASVLPGLLRQADPOHVIEYSALVTVLNEYERALDVAAEPKHERQHQRAIRKNIT 2647  
QY 354 -----A-----I--K----- 356  
Db 2648 ETLVSLRVHTVDDIOIAAALAAQCMGPRSELVCRSLKOTLHKLEAMMLILOAETTATGV 2707  
QY 357 -----WD-----PT--G-----NL----- 363  
Db 2708 TPTAIGDSILNITGDLIHLASSDVRAPOFSELGAESPMSWASQANLTSALMRLMRSR 2767  
QY 364 -----LA-----S-----C-----SD----- 369  
Db 2768 VLNEEPLTLAGEEIVAGKESDRSLCYGAPGCGHFSIPEAFSGALANLSDVVQLIF 2827  
QY 370 --D-----M--T-----L-----KI-----W----- 376  
Db 2828 LVDSNPPFPGYISNYTVSTKVSMAPQOAGAPIERLASERAITVKVPNNSDMAARGH 2887  
QY 377 -----SM-----KO-----D----- 381  
Db 2888 RSSANSVSVVQFQASVAGVWTLSDSNPAAGHLQNLQNTLLDGHVLSPEPEYLAIVLH 2947  
QY 382 -----NC-----V-----H-----DLQA--H--N-----K--E--I 394

Db 2948 SBPRPNEHNCASRRIRPESLOQADHRPYTFTISGSRD--PAGSYHLNLSHFWRFSALQV 3006  
QY 395 ----YT-----IKW-----SP-----T--G----- 403  
Db 3007 SVGLYTLSCQYFSEEDMV--WRTEGLLPLEETSPRQAVCLTRHLTATFASGLFVPPSHVRV 3065  
QY 404 --P----- 404  
Db 3066 FPEPTADVNIYVMLTCAVCLVTYVMAAILHKLQDLQDASGRRAIPFCGGRGKFKYBILVK 3125  
QY 405 -----GT-----N----- 407  
Db 3126 TCMGRSGTTHAHVIMLYGVDSRSGHRLDGDAPHRNSLDIPRIATHPSLGSVMKIRVW 3185  
QY 408 --N--P-----N-----AN--L--M--L--AS--AS-- 419  
Db 3186 HDNKGSLPAWFIQHVIVRLQOTARSFAFFLVNDWLSVETANGGLVEKEVLAASDAALLRF 3245  
QY 420 -----FD-----S--T--V--R-----L-----W-----D-- 428  
Db 3246 RRLVLAELQRFQKHIMLSIWDPRPSRFTRIQATCCVLLICLFLGANAVMYGAVGDS 3305  
QY 429 -----VD-----R-----G----- 432  
Db 3306 AYSTGHVSRSLPSVDTVAVGLVSSVYVYVYLAAILFLFRMSRSKVAGSPSPAGQVQL 3365  
QY 433 -I--C--I-----H-----TLT----- 439  
Db 3366 DIDCLDSSVLDSFLTFPSGLHAEAFVGMQKSDLFDDSKSLVCWPSGGTJLWPDLLSD 3425  
QY 440 -----K----- 440  
Db 3426 PSIVGNLRLQARGQAGHGLGPEEDGFSLASPYSPAKSPFSADEDLIQQVLAEVSSPAP 3485  
QY 441 --H-----O-----E----- 443  
Db 3486 TQDTHMETDLLSSSTPGTEKTETLALQRLGBLPPSPGLNWEQQAARLSRTGLVEGLR 3545  
QY 444 --P-----V--Y-----SVA----- 449  
Db 3546 KRLLPAMCASLAHGLSLLLVAVAVAGVWGASFPFGVSVAMLLSSSASFLASFLGWEP 3605  
QY 450 -----PS-----PD----- 453  
Db 3606 KYLLEALYFSLVAKRLHPDEDDTLVESPAVTPVSARVPRVRPHGFALFLAKEEARKVR 3665  
QY 454 --G--R--Y-----LAS-----G-----S--F----- 462  
Db 3666 LHGMLRSLIVYMLFLVLTLLASYGDSACHGAVYRLQSAIKQELHSRAFLAITRSEELWPW 3725  
QY 463 -----DK-----C-----V----- 466  
Db 3726 MAHVLLPYVHGNQSSPELGPRLRQVRLQALYPOPPGPRVHTCSAAGFSTSDYDVGWE 3785  
QY 467 --H-----I-----W--N-- 470  
Db 3786 SPHNGSGTWAYAPDLLGAWNGSCAVYDSGGVQELGSLBESDRDLRFLQLHNLNDR 3845  
QY 471 -----TQ-----T-----G-----AL----- 476  
Db 3846 SRAVLELTRYSPAVGLHAAVTLRLLEFPAGRALAALSVRFPALRRLSAGLSLPLTTSVC 3905  
QY 477 -----VH-----SY-----R-----G-----T-- 483  
Db 3906 LLLFAVHFAVBARTWHREGWRVRLGAWARWLLVALTAATVALVLAOLGAADROWTRF 3965  
QY 484 --G-----G-----IF-----E-- 488  
Db 3966 VGRPRRFTSPQVAHVSSAARGLAASLLFLLLVKAAQHVRFVRQWSVFGKTLRALPEL 4025  
QY 489 -----V----- 489  
Db 4026 LGVTGLGVVAVYAQLAILLVSSCVDSLWSVAQALLVLCPGTGLTCLPABSWHLSPLL 4085

QY 490 C-----WNA-----A-G-----D----- 496  
Db 4086 CUGLWALRLMGALRGALVILRWYHALRGELYRPAWEPQDYEMVELFLRLRLMMGLSKV 4145  
QY 497 -----KV-----G-----A-S-AS-----DG-SV----- 507  
Db 4146 KEFRHKVRFEGMEPLPSRSGKSVDPDPPPSAGSDASHPTSSSQDGLSVSLGRJGT 4205  
QY 508 -C-----V-----L-----D-----L--RK 514  
Db 4206 RCEPEPSRLQAVFEALLTQFDRLNQATEDVYQLEQQLHLSLQGR 4249

## RESULT 14

US-09-052-262-8  
; Sequence 8, Application US/09052262  
; Patent No. 6656681  
; GENERAL INFORMATION:  
; APPLICANT: Harris et al.  
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE  
; TITLE OF INVENTION: AND USES THEREOF  
; NUMBER OF SEQUENCES: 23  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff, Ltd.  
; STREET: One Financial Center  
; CITY: Boston  
; STATE: MA  
; COUNTRY: US  
; ZIP: 02111  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk, 3.50 inch  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: WordPerfect 6.1  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/052,262  
; FILING DATE: Concurrently herewith  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/422,582  
; FILING DATE: 14-April-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9507766.5  
; FILING DATE: 13-APR-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9411900.5  
; FILING DATE: 14-JUN-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB PCT/GB94/02822  
; FILING DATE: 23-DEC-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9326470.3  
; FILING DATE: 24-DEC-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Williams, Ph.D., Kathleen M.  
; REGISTRATION NUMBER: 34,380  
; REFERENCE/DOCKET NUMBER: 3265/74118  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617) 345-9100  
; TELEFAX: (617) 345-9111  
; INFORMATION FOR SEQ ID NO: 8:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 4302 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-09-052-262-8

Query Match 69.5%; Score 2548.4; DB 4; Length 4302;  
Best Local Similarity 9.8%; Pred. NO. 1e-31; Gaps 357;  
Matches 412; Conservative 83; Mismatches 13; Indels 3676;

QY 1 MSIS-----S-----D-----EVN-----F----- 10

Db 72 LDVSHNLRALDVGLANLSALAEIDI SNKISTLEEIGIFANFLNLSEINLSGNPFECDC 131  
QY 11 -LVY--RY-----L-Q-----E-----SQ- 20  
Db 132 GLAWLPRWAEEOQVRVQPEATCAGPSLAGOPLLGLLDSCGGEYVACL PDSNST 191  
QY 21 ----FS-H-----SAF----- 26  
Db 192 VAAVSFAAHEGLLOPEACSAFCFTGQGLAALSQGWCLCGAAQPPSSASFACLSICSP 251  
QY 27 -----T-----F-----G-----I----- 30  
Db 252 PPPAPTCKGPTLLQHVFPASPGATLVGPHGPLASGQLAAFHIAAFLPVTATRWDFGDS 311  
QY 31 -E-----SH-----I-----S-QSN- 38  
Db 312 AEVDAAGPAASHRYVLPGRYHYTAVLALGAGSALLGTDVQVEAAPAALELVCPSSVQSD 371  
QY 39 ----I-N-G-----AL-----V-----P----- 46  
Db 372 SLDSLQNRGGSGLEAAYSI VALGEBEPARAVHPLCFSDTEIFPGNGHCYRLVVEKAAWLQ 431  
QY 47 -----AA-----LIS-----I-----Q-K-----G-----L----- 57  
Db 432 AQEQCAWAGAAALAMVDSPAVQRFVLSRVTRSLDVWIGFSTVQGVVEGPAQGEAFSL 491  
QY 58 -Q-----YV-E-----AE-----V-----S- 65  
Db 492 CQNLFGPEHPATAHCVRLGPTGMCNTDLCAPHSYVCELOPGGVPQDAENLLVGAPSG 551  
QY 66 ----IN-----BDG-----TL-----F----- 73  
Db 552 DLQGLTPLAQDGLSAPHEPVVVFPGRLSLRBAFLTTAEFGTQELRRPAQLRQVY 611  
QY 74 -----D-----GR-P-I----- 78  
Db 612 LLSTAGTPENGSEPSRSPDNRTQLAPACMPGGRWCPGANICLPDASCHPOACANGCTS 671  
QY 79 -----E-S-----LSLI-----DAV-M-P-DVV----- 92  
Db 672 GPGCPAYALWREFLSVAAGPPAQYS-VTLHGQD-VLMLPGDLLVGLQHDAGPGALLHC 729  
QY 93 -----Q-----T-----R-----Q----- 96  
Db 730 SPAPGHPGQAPYLSANASSWLPPLPALEGTWACPCALRLIAATEQLTVLLGLRPNPG 789  
QY 97 -----QA-----Y-----RD-KL-----A-----Q- 105  
Db 790 LRMPGRYEVRAEYGVNGVSRHNLSCSPDVVSPVAGLRVIYPAPRDRGLYVPTNGSALVLQ 849  
QY 106 ----- 107  
Db 850 DSGANATATARPGGSVARFENVCALVATFVPGCPWETNDTLFSVVALPWLSEGHVV 909  
QY 108 -----A-----A-----A-----A----- 113  
Db 910 DVVENSASRANLSLRVTAEEPIGGLRATPSPEARVLQGVLRVSPVWAGSDMVFRTI 969  
QY 114 -----AA-----A-----AT-----N-----Q----- 121  
Db 970 NDQSLTFQNVFNVIYQSAAVEFKLSLTASNHVSNVTNVNVTVERNNRQGLQVSTVPA 1029  
QY 122 -----G-----SA-K-----N-----GE-----N-----TA----- 131  
Db 1030 VLSFNATLVLTGVLVDSAVEVAFLWNFGDGEQALHQFPQPPYNESFPVDPSPVAQLVEH 1089  
QY 132 N-----GE-----EN-----G-----A-----H 139  
Db 1090 NVMHTVAAAPCEYLLTVLASNAFENLAQOVFVSVRASLPVAVGSDGLVAGRPVTFYPH 1149  
QY 140 -----T-----I-----ANNHT----- 146

Db 1150 PLSPGGLVLTWDFGDSVLTQSPAAHNTYASRGTYHVRLEVNNTVSGAAAQADRVF 1209  
QY 147 ---DM-M-E---V---D-GD-----154  
Db 1210 BELGLSDMSLAVEGAPVWVSAAVQTGDNITWTDFMDGTVLSCPEATVEHVYLRQN 1269  
QY 155 ---VE---IPS-----NKA-----162  
Db 1270 CTVTGGAASPAGHLARSLHVLVFLVLRVEPAACIPTQPDARLTAYVTGNPAHYLFDWT 1329  
QY 163 ---V---L-R-G---H---ESEV-----172  
Db 1330 FGDSSNTVRCPTVTHNFTSRGTPLALVLSRVNRHAYFTSICVEPEVGNVLQPER 1389  
QY 173 -FI-----CAM-----NP-----179  
Db 1390 QFVLGDEAWLVACAPPPPYRYTDFGTEEAAPTARGPEVTFIYRDDPGSYLVTVTASN 1449  
QY 180 -VS---D-L---LV-S---GS-----GD-----190  
Db 1450 NISAANDSALVEQEPVLVTSIKVNGSLGLELQOPLYFSAVGRGRPASVYLDLGDGWL 1509  
QY 191 ---S---TARI-WN-L-SE---NST---SG-----P-----206  
Db 1510 GPEVTHAYNSTGDTVRVAGMNEVSRSEAWLNTYVRRVGLVYNASRTVPLNGSVSFS 1569  
QY 207 T-----QL--VL--RHC-----I--RE-G---G-QD-----221  
Db 1570 TSLEAGSDVRYSWLDCR-CTPIPGPTISYTRFSVGTNIIIVTAENEVGSQDSIFVVV 1628  
QY 222 ---V---PSN-----XD---V---T-----SL-----231  
Db 1629 LQLEGLVQVGGGRYFPTNHTVQLQAVRDGTVNSYSWTAWDRGPALAGSGKGFSLTVL 1688  
QY 232 ---D---W-----N-----S---235  
Db 1689 EAGTYHVQLRATNMLGSAMADCTMDFEVPVGLMTVTASNPAAVNTSVTLASLAGSGV 1748  
QY 236 ---EG---T-L-LAT---GS-----YD-----246  
Db 1749 VYTWSLEGLSWETSEPTTHSTPGLHLVTVTGNPLGSANATVEVDVQVPVSLSR 1808  
QY 247 ---G-F-A---R---I---W-----TK---D-G-----NL 258  
Db 1809 ASBPGGSFVAAGSSVPFGQLATGTVNWCWAVPGGSSKRGPHVTWVPDAGTFSIRLN- 1867  
QY 259 AS-----T-----L-----GQ-H-----K---G---267  
Db 1868 ASNAVSWVSATYNTLTAEEPIVGLVLWASSKVAVPQGLVHFQILLAAAGSAVTFRLQVGGAN 1927  
QY 268 P-I---F-AL---K---W-----N-KKG---278  
Db 1928 PEVLPGFRFHSPPRGDHHVSVRGKQNVSWAQOQRIVVLEAVSGLQMPNCCPEGIATG 1987  
QY 279 ---NF-----ILS-----AG-----285  
Db 1988 TERNFTARVQGRSAVAYFSLQKVGDSLVLSGRDVTYTPVAAGLLEIQVRAFNALG 2047  
QY 286 ---V-D---K---T-----T-----II-WD-----294  
Db 2048 SENRTLVEQDAVQVVALQSGPCTNRSQAPEAATSPSPRVAYHWFDEGDSFGQDDE 2107  
QY 295 --A-HT---G---E-A---KO---O-----FP-----305  
Db 2108 PRAEHSYLRPGDYRVQVNASNLVSFFVAQATVTVQLACREPDVVVLPQLVLMRRSQRN 2167  
QY 306 F---H---SA-----P---AL---DV-----314  
Db 2168 YLEAHVDLRDCVTYQTEYRWEVYRTASCQRPGRPARVALPGVDVSRPLVLPALPVGH 2227  
QY 315 -----D-----W-----QS---N---319  
Db 2228 YCFVFWVSFGDTLQSIQANVTVAPELVPITIEGGSYRWSDTDLVLDGSSSDPNLE 2287

QY 320 ---N-----TF-----322  
Db 2288 DGDQTPLSFHWACVASTQREAGGCALNPGRGSSVTIIPRERLAAGVEYTFSLTVWKAGR 2347  
QY 323 ---A-----SC-----325  
Db 2348 KEEATNQTVLIRSGRPVIVLSVCVSCKAQAVYEVSRSSVYVLEGRCLNCSGSKRGWAA 2407  
QY 326 ---S-T-D---M-----C-I-----331  
Db 2408 RTFSNKTLDVDTTSTGSAHRLVLRGVLRDGEYTFTLTVLGRSGEBCASIRLSP 2467  
QY 332 ---H-----V-----C-334  
Db 2468 NRPLGSGCRLPPLGAVHALTTKVHFECTGMDAEDAGAPLVVALLRRRCQGHCEFCV 2527  
QY 335 -K-----L-----G-----QD-----R-----P-----I-342  
Db 2528 YKGSLSYGAVLPPGFRPHFEVGLAVVQDOLGAADVVALNRSLAITLPEPNGSATGLTVM 2587  
QY 343 ---KT---PQG---H---T-NE---V-----N---353  
Db 2588 LHGLTASVLPGLLRQADPOHVEYLSALVTVLNEYERALDVAAEPKHERQRAQIRKNIT 2647  
QY 354 -----A-----I---K-----356  
Db 2648 ETLVSLRVHTVDDIQIIAAALAQCMPSRELVCRSCLKOTLHKLEAMMLILQAEATTAGTV 2707  
QY 357 ---WD-----PT-G---NL-----363  
Db 2708 TPTAIGDSILNITGDLIHLASSDVRAPQPSSELGAESPMSRMVASQAYNLTSALMRILMSR 2767  
QY 364 ---LA-----S-----C-----SD-----369  
Db 2768 VLNEBPLTLAGEBEIWAQGRSDPRLLCYGAPGCGCHFSIPEAFSGALANLSDVVQLIP 2827  
QY 370 ---D---M---T-----L-----KI---W-----376  
Db 2828 LVDSNPFPFGYISNYTVSTKVASMAFQOQAQAIPIERLASERAITVKVPNNSDMAARGH 2887  
QY 377 ---SM---KQ-----D-----381  
Db 2888 RSSANSANVVVQPOASVGAVVTLDSNPAAGLHLQLNLTLLDGHYLSBEPEPYLAVYLH 2947  
QY 382 ---NC---V-----H-----DLQA---H-N---K---EI 394  
Db 2948 SEPRNEHNCASRRIRPESLQGAHDPYTFIFSPGSRD-PAGSYHLNLSHFRNSALQV 3006  
QY 395 ---YT---IKW-----SP-----T-G-----403  
Db 3007 SVGLYTSLOQYFSEEDMV-WRTEGLLPLEETSPRQAVCLTRHLTAFASLFPVPPSHRVFV 3065  
QY 404 ---P-----404  
Db 3066 FPEPTADVNIYVMLTCAVCLVTVMMAAILHLKLDLDASGRRAIPFCQGRGFKYBILVK 3125  
QY 405 ---GT---N-----407  
Db 3126 TGWRGSGTTHVGMVYGVDSRSGHRHLDGDRAFHRNSLDIFRIATPHSLGSLVKIRVW 3185  
QY 408 ---N---P-----N-----AN-L-M--L-AS-AS---419  
Db 3186 HDNKGSLSPAWFLQHVLRDLQARSFAFFLVNDWLSVETEANGGLVEKEVLAASDAALLRF 3245  
QY 420 ---FD-----S-T-V-R-----L-----W-----D-428  
Db 3246 RRLVLAELQRFKDHIMLSIWDPRPSRFRTRIQRATCCVLLICLFLGANAVMYGAVGDS 3305  
QY 429 ---VD-----R-----G-----432  
Db 3306 AYSTGHVSRLSPSLVDTVAVGLVSSVVVYVYLAILFLFRMSRKVAGSPPTPAGQOVL 3365

```
QY 433 -I--C--I--H-----TLT-----439
Db 3366 DIDSCLDSSVLDSSFLTFSGLHAEAFVGQKSLDFLDDSKSLVCWPSGEGTLLSWPDLSSD 3425
QY 440 -----K-----440
Db 3426 PSIVGNLRQLARGQAGHGLGPEEDGFSGLASPYSPAKSPFSADEDLIQVLAARGVSSPAP 3485
QY 441 -----Q-----E---443
Db 3486 TDTHMETDLLSSLTGPKETETLALQRLGELGPPSPGLNWPQQAARLSRTGLVEGLR 3545
QY 444 -----V--Y-----SVA-----449
Db 3546 KRLLPAWCASLAHGLSLLLVAVAVAGVWGASFPFGVSVANLLSSASFLASFLGWEPL 3605
QY 450 -----PD-----453
Db 3606 KVLEALYFSLVAKRLHPDEDDTLVESPATVPVSARVVRPHGFALFLAKEEARKVKR 3665
QY 454 -G-R--Y-----LAS-----G-----S-F-----462
Db 3666 LHGMLRSLVYMLFLVTLVLLASYGDASCHGHAYRLQSAIKQELHSRAFLAITRSEBELWPW 3725
QY 463 -----DK-----C-----V---466
Db 3726 MAHVLLPYVHGNOSSBELGPRRLRQVRLQALYDPDPGPRVHTCSAAGGFSTSDYDVGWE 3785
QY 467 --H-----I-----W--N--470
Db 3786 SPHNGSTWAYSAPDLLGAWSGSCAVYDGGVQELGLSLEBSRDLRLPLQLHNLNDR 3845
QY 471 -----TO-----T-----G-----AL-----476
Db 3846 SRAVLELTRYSPAVGLHAAVTLRLFPFPAAGRALAALSVERPFALRRLSAGLSPLLTVC 3905
QY 477 -----VH-----SY-----R-----G-----T---483
Db 3906 LLLFAVFAVAEARTHREGWRVRLGAWARMLLVALTALVRLAQLGAADROWTRF 3965
QY 484 --G-----G-----IF-----E-488
Db 3966 VRGPRRFTSDQAVHSSAAGLAASLLFLLLVKAAQHVRFVQWSVFGKTLRALPEL 4025
QY 489 -----V-----489
Db 4026 LGVTGLVILGVAYQAAILLVSSCVDSLWSVAQLLVLCPGTGLTCLCPAESWHLSPLL 4085
QY 490 C-----WNA-----A--G-----D-----496
Db 4086 CVGLWALRLMGLALGLGAVILRWYHALRGELYRPAPWEPQDYEMVELFLRLRLMGLSKV 4145
QY 497 -----KV-----G-----A-S-AS-----DG-SV-----507
Db 4146 KEFRHKVREGMEPLPSRSRSGSKVSPDVPFPSPAGSDASHPTSSSQLDGLSVSLGRIGT 4205
QY 508 -C-----V-----L-----D-----L--RK 514
Db 4206 RCEPEPSRLQAVFEALTQFRLNQATEDVYQLEQLHSLQGR 4249
```

## RESULT 15

US-09-052-469-6

; Sequence 6, Application US/09052469

; Patent No. 6380360

; GENERAL INFORMATION:

; APPLICANT: Harris et al.

; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE 1 GENE

; TITLE OF INVENTION: AND USES THEREOF

; NUMBER OF SEQUENCES: 23

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Banner &amp; Witcoff, Ltd.

; STREET: One Financial Center

```
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk, 3.50 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; FILING DATE: Concurrently herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/422,582
; FILING DATE: 14-APRIL-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9507766.5
; FILING DATE: 13-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9411900.5
; FILING DATE: 14-JUN-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB PCT/GB94/02822
; FILING DATE: 23-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9326470.3
; FILING DATE: 24-DEC-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Ph.D., Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3265/74165
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 345-9100
; TELEFAX: (617) 345-9111
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4339 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-052-469-6
```

Query Match 69.5%; Score 2548.4; DB 3; Length 4339;

Best Local Similarity 9.8%; Pred. No. 1.1e-31;

Matches 412; Conservative 83; Mismatches 13; Indels 3676; Gaps 357;

```
QY 1 MSIS-----S-----D-----EWN-----F---10
; : : : : : : : : : : : : : : : : : :
Db 29 LDVSHNLLRALDVGLLANLSALAEALDISNNKISTLEEGIFANLFLNLSGNPFECDC 88
QY 11 -LVY--RY-----L--Q-----E-----SG-20
; : : : : : : : : : : : : : : : : : :
Db 89 GLAWLPRWABEQVRVQPEAATCAGPGSLAGQPLLIGIPLLDGCGGEEYVACLPDNSSGT 148
QY 21 ----PS--H-----SAF-----E-----26
; : : : : : : : : : : : : : : : : : :
Db 149 VAAVFSAAHEGLQPEACSAFCFTGQGLAALSEQWCLCGAAQPSASFACLSLCSGP 208
QY 27 -----T-----F-----G-----I-----30
; : : : : : : : : : : : : : : : : : :
Db 209 PPPAPTCTRGPTLLQHVFPASPATLVGPHGLASGLAAPHIAAPLPVTATRWDFGDGS 268
QY 31 -E-----SH-----I---S-QSN-38
; : : : : : : : : : : : : : : : : : :
Db 269 AEVDAAGPAASHRYVLPGRHYVTAVIALGAGSALLGTDVQVEAAPAALELVCFSSVQSD 328
QY 39 -----I-N-----G-----AL-----V-----P-----46
; : : : : : : : : : : : : : : : : : :
Db 329 SLDLSIQNRGGSGLEAAYSIVALGEERPARAVHPLCFSDTEIPFGNGHCYRLVVEKAAWLQ 388
QY 47 -----AA-----LIS-----I-----IQ--K-----G-----L--57
; : : : : : : : : : : : : : : : : : :
Db 389 AOEQQAWAGAALAMVDSFAVQRFVLSRVTRSLDVMWIGFSTVQGVGVPAPQGEAFSLES 448
```



QY 58 Q-----YV-E-----AE-V--S- 65  
Db 449 CQNLPEHPATAEHCVRGLPTGCWNTDLCSPHYSVCELOPGFVQDAENLLVCAPSG 508  
QY 66 ---IN-EDG-----TL--F----- 73  
Db 509 DLQPLTLPLAQDGLSAPHPEVVMVFPGLRLSREAFLLTAETGTOELRRPAQLRLQVVR 568  
QY 74 -----D-----CR--P--I----- 78  
Db 569 LLSTAGTPENGSEPSRPNRTQLAPACMPGRCWCPGANICLPLDASCHPQACANGCTS 628  
QY 79 -----E--S-----LSLI-----DAV-M-P-DV- 92  
Db 629 GPGLPGAPYALWREPLFSVAAGPPAQYS-VTLHQD-VLMLPGDVLGLQHDAGPGALLHC 686  
QY 93 -----Q--T-----R-----Q----- 96  
Db 687 SPAPGHGPOAPYLSANASSWPLPPLPAQLEGTWACPACALRLLAATEQLTVLLGLRPNG 746  
QY 97 -----QA-----Y--RD-KL-----A--Q- 105  
Db 747 LRMGRYEVRAEVGNGVSRHNLSCSFDVSPVAGLRVVPAPRDRGLYPTNGSALVLQV 806  
QY 106 -----AA-----AT--N-Q-Q-----OH-- 107  
Db 807 DSGANATATARWPGGSVARSPPENVCALVATFVPGCPWETNDTLFSVVALPMLSEGEHV 866  
QY 108 -----A--A-----A-----A----- 113  
Db 867 DVVENSASRANLSLRTAEPIICGLRATPSPPEARVLQGLVLRYSVPVBEAGSDMVFRTI 926  
QY 114 -----AA-----AT--N-Q-Q----- 121  
Db 927 NDQSLTFQNVFNVIYQGAFLKSLTASNHVSNTVNVNVTVERMRMQGLQVSTVPA 986  
QY 122 -----G--SA-K--N--GE-----N--TA----- 131  
Db 987 VLSPNATLVTGVLVDSAVEAFLNFGDGEALHQFPNPNESFPVDPSPVAQVLVEH 1046  
QY 132 N-----GE-----EN-----G-----A-----H 139  
Db 1047 NVMHYAAPEYLLTVLASNAFENLQQVPVSVRASLPSVAVGSDGLVAGRPVTFYPH 1106  
QY 140 -----T-----I-----ANNHT----- 146  
Db 1107 PLPSPGGVLYTWDGDSVLTQSPAAHNTYASRGTYHVRLEVNNTVSGAAAQADRVF 1166  
QY 147 -----DM-M--E-----V-----D-GD----- 154  
Db 1167 EELRGLSDMSLAVEGAPVUVSAAVQTGDNITWTFMDGDGTVLGSPEATVEHVYLRQN 1226  
QY 155 -----VE--IPS-----NKA----- 162  
Db 1227 CTVTGGAASPAHLSRLHVLVFLVLRPEAPACIPTQDARLTAYVTGNPAHYLFDMT 1286  
QY 163 -----V-----V--L-R-G-----H-----ESEV----- 172  
Db 1287 FGDSSNTTVRGCTPTVTHNFTSGTFLALVLSSRVNRAHYFTSICVEPEVGNVTLOPER 1346  
QY 173 -FI-----CAW-----NP----- 179  
Db 1347 QFVLGDEAWLVACAMPPPPRYTWDFGTGEEAAPTRARGPEVTFIYRDFGSYLVTVTASN 1406  
QY 180 -VS--D--L-----LV-S--GS-----GD-- 190  
Db 1407 NISAANDSALVEQEPVLVTSIKVNSLGLLEQLQPYLFSVAGRGPRASLYLDLGDGGMLE 1466  
QY 191 -----S-----TARI--WN-L--SE--NST-----SG-----P----- 206  
Db 1467 GPEVTHAYNSTGDFTVRAGWNEVSRSEAWLNVTVRKRRVGLVYNASRTVWPLNGSVSFS 1526

QY 207 T-----QL--VL--RHC-----I-----RE-G-----G--QD----- 221  
Db 1527 TSLEAGSDVRYSWLVCDDR-CTPIPGGPTISYFRSVGTNIIIVTAENEVGSQADSIFVVV 1585  
QY 222 -----V-----PSN-----KD--V--T-----SL-- 231  
Db 1586 LQLIIEGLQVVGGRYFPPTNHTVQLQAVVRDGTNVSWTAWDRGPALAGSGKGFSLTVL 1645  
QY 232 -----D-----D-----W-----N-----S-- 235  
Db 1646 EAGTYHVQRATNMLGSAWADCTMDFVEBPVGMVMVTASNPAAVNTSVLSAELAGSGV 1705  
QY 236 -----EG-----T--L--LAT-----GS-----YD----- 246  
Db 1706 VYTWISLEGLSWETSEPTTHSPPTPGLHLVTMTAGNPLGSANATVEVDVQVPVSGLSIR 1765  
QY 247 -----G-P--A-----R-----I--W-----TK--D-G-----NL 258  
Db 1766 ASEPGGSFVAAGSSVPFMGQLATGTNVSMCWAVPGGSSKRGRPHVTMVPFDAGTFSIRLN- 1824  
QY 259 AS-----T-----L-----GQ--H-----K-----G-- 267  
Db 1825 ASNAVSWSATNLTABEPIVGLVWASSKVAPGOLVHFQILLAAAGSAVTFRLQVGGAN 1884  
QY 268 P-I-----F--AL-----K--W-----N--KKG----- 278  
Db 1885 PEVLPGPRFHSHPFVRGDHVVSVRGKHNHVSMAQAQVRIVVLEAVSGLQMPNCCPEGIATG 1944  
QY 279 ---NF-----ILS-----AG----- 285  
Db 1945 TERNFTARVQGRSVAYAWYFSLQVQGSVLSDVLTGRDVTYTPVAAGLLEIQVRAFNALG 2004  
QY 286 -----V-D-----K--T-----T-----II--WD----- 294  
Db 2005 SNRTLVLVEQDAVQVVALQSGFCFTNRSAQFEAATSPSPRRVAVYHWDGDSGPGQDTE 2064  
QY 295 ---A-HT-----G--E--A-----KQ-----Q-----FP----- 305  
Db 2065 PRAEHSYLRPGDYRVQVNASNLVSFFVAQTVTQVLACREPEVDVVLQVLMMRSQRN 2124  
QY 306 F--H-----SA-----P--AL--DV----- 314  
Db 2125 YLEAHVLDLDCVTYQTEYRWEVYRTASQCORPARVALPGVDVSRPLVLRPLALPVGH 2184  
QY 315 -----D-----D-----W-----QS--N-- 319  
Db 2185 YCFVFWSPGDTPLTQSIQANVTVAPERLVPIIEGGSYRVWSDTRDLVLDGSESYDNLE 2244  
QY 320 -----N-----N-----TF----- 322  
Db 2245 DGDQTPLSFWACVASTQREAGGCALNFGPGSSVTIIPRERLAAGVEYTFSLTVWKAGR 2304  
QY 323 ---A-----SC----- 325  
Db 2305 KEEATNQTILIRSGRPVIVSLEVCSCQAQAVYEVSRSSVYVLEGRCLNCSGSKRGWAA 2364  
QY 326 ---S--T--D-----M-----C--I----- 331  
Db 2365 RTFSNKTLLVDETTTSGAGMRLVLRGVLDRDGBGYTFTLVLRGSGBEEGCASIRLSP 2424  
QY 332 -----H-----V-----C-- 334  
Db 2425 NRPPILGSGCRPLPLGNVHALTTKVHFECTGMDAEDAGAPLVYALLRRRCROGHCEFCV 2484  
QY 335 -K-----L-----G--QD-----R-----P-----I-- 342  
Db 2485 YKGLSSYGAVLPFGPRPHFVEGLAVVQDQLGAAVVALNRSLAITLPEPNGSATGLTW 2544  
QY 343 ---KT--FQ-----H-----T--NE-----V-----N-- 353  
Db 2545 LHGLTASVLPGLLRDADPOHVIEYSLALVTVLNEVERALDVAEPKHQHQRAQIRKNIT 2604  
QY 354 -----A-----I--K----- 356

```
Db 2605 ETLVSLRVHTYDDIQIAAALACQMGSPRELVCRLKQTLHLKLEAMMLILQAEITAGTV 2664
Qy 357 -----WD-----PT-G-----NL----- 363
Db 2665 TPTAIGDSILNITGDLIHASSDVRAPQPSBELGAESPSPRMVASQAYNLTSALMRLMSR 2724
Qy 364 -----LA-----S-----C-----SP----- 369
Db 2725 VLNEEPLTAGEEIVAOGRSDPSRLCYGAPGPGCHFISPEAFSGALANLSDVQLIF 2784
Qy 370 -----D-----M-----T-----L-----KI-----W----- 376
Db 2785 LVDNPPFPFYISNYTVSTKVASNAFQOAGAIPIERLASERAITVKVPNNSDWAARGH 2844
Qy 377 -----SM-----KO-----D----- 381
Db 2845 RSSANSANSVVVQPQASGVAVTLDSSNPAAGHLQNLNLTLDGHYLSSEPEPYLAVLH 2904
Qy 382 -----NC-----V-----H-----DLQA-----H-N-----K-----EI 394
Db 2905 SEPRNEHCASRRIRPESLOGADHRPYTFIFSGSRD-PAGSYHNLNLSHFWSALQV 2963
Qy 395 -----YT-----IKW-----SP-----T-G----- 403
Db 2964 SVGLYTSLCQYFSEEDMV-WRTEGLLPLEETSPQAVCLTRHLTAFCASLFVPPSHVRFV 3022
Qy 404 -----P----- 404
Db 3023 FPEPTADVNYIVMLTCAVCLVYVMMAAILKLDQDASGRATPFCQGRFKYEILVK 3082
Qy 405 -----GT-----N----- 407
Db 3083 TGWRGSGTTHAHGIMLYGVDSRCHRHLDGDRAFHRNSLDIFRIATPHSLGSLVWKIRW 3142
Qy 408 -----N-----P-----N-----AN--L--M--L--AS--AS-- 419
Db 3143 HDNKGSLPAWFLQHVIVRDLQTARSAFFLVNDMLSVETEANGGLVEKEVLAASDAALLRF 3202
Qy 420 -----FD-----S--T-V-R-----L--W-----D-- 428
Db 3203 RRLVABLQRFKDKHILSIWDRPPRSRFRTRIQATCCVLLICLFLGANAVVYAGVDS 3262
Qy 429 -----VD-----H-----R-----G----- 432
Db 3263 AYSTGHVSLSPSLVDIVAVGLVSSVVVYPVYLAILFLFRMSRKVAGSPSPFAGQOVL 3322
Qy 433 -----I--C-----I-----H-----TLT----- 439
Db 3323 DIDSCLDSSVLDSFLTFTGSLHAEAFVGMKSDLFLDSDSKLVCWPSGEGTSLWPDLSD 3382
Qy 440 -----K----- 440
Db 3383 PSIVGSNRLQARQAGHLGPEEDGFSLASPSPAKSFSASDEDLIQVLAEGVSSPAP 3442
Qy 441 -----H-----O-----E----- 443
Db 3443 TQDTHMETDLLSSLSSTPEKTELTALQRLGELGPPSPGLNWEQPQAARLSRTGLVEGLR 3502
Qy 444 -----P-----V--Y-----SVA----- 449
Db 3503 KRLLPACASLAHGLSLLLVAVAVAGVGMVGSFPFGVSVALLSSSASFLASFLGWEPL 3562
Qy 450 -----FS-----PD----- 453
Db 3563 KVLLEALYFLSVAKRLHPDSDDTLVESPAVTPVSARVVRPVRPHGFALFLAKEEARKVR 3622
Qy 454 -----G--R-----Y-----LAS-----G-----S--F----- 462
Db 3623 LHGMRLSLVYMLFLLVTLASYGDASCHGHAYRLQSAIKQELHSRAFLAITRSEELWPW 3682
Qy 463 -----DK-----C-----V----- 466
```

Search completed: January 3, 2005, 15:30:42  
Job time : 93.6667 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:26:14 ; Search time 81 Seconds

(without alignments)  
2282.717 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 3669

Sequence: 1 MSISDEVNVLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 1599051 seqs, 359727711 residues

Total number of hits satisfying chosen parameters: 1599051

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10A\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3350.2	91.3	577	17	US-10-723-860-836
2	3260.6	88.9	542	15	US-10-264-049-2892
3	2716	74.0	395	16	US-10-363-829-446
4	2631.9	71.7	5635	15	US-10-451-168-78
5	2627.8	71.6	5636	14	US-10-032-189-128
6	2627.8	71.6	5636	14	US-10-120-801-72
7	2627.8	71.6	5636	14	US-10-023-634-93
8	2627.8	71.6	5636	16	US-10-408-765A-1895
9	2612.9	71.2	4599	15	US-10-464-368-69
10	2612.9	71.2	4599	15	US-10-464-368-70
11	2612.1	71.2	5175	14	US-10-120-801-74
12	2612.1	71.2	5175	14	US-10-369-493-6859
13	2612.1	71.2	5175	14	US-10-369-493-6861

14	2610.8	71.2	5198	14	US-10-120-801-75	Sequence 75, Appl
15	2610.8	71.2	5198	14	US-10-369-493-6858	Sequence 6858, Ap
16	2610.8	71.2	5198	14	US-10-369-493-6860	Sequence 6860, Ap
17	2610.8	71.2	5198	15	US-10-346-863-34	Sequence 34, Appl
18	2594.4	70.7	4599	17	US-10-479-875-4	Sequence 4, Appl
19	2593.4	70.7	4636	9	US-09-835-996A-33	Sequence 33, Appl
20	2592.2	70.7	26926	9	US-09-759-508B-2	Sequence 2, Appl
21	2592.2	70.7	26926	17	US-10-723-860-133	Sequence 133, App
22	2588.8	70.6	4183	13	US-10-087-192-672	Sequence 672, App
23	2588.8	70.6	4183	16	US-10-741-601-332	Sequence 332, App
24	2588.8	70.5	4563	15	US-10-276-774-1723	Sequence 1723, Ap
25	2587.7	70.5	4545	9	US-09-873-403-2	Sequence 2, Appl
26	2587.7	70.5	4545	11	US-09-750-972-2	Sequence 2, Appl
27	2586.1	70.5	4544	15	US-10-464-368-68	Sequence 68, Appl
28	2586.1	70.5	4544	17	US-10-479-875-6	Sequence 6, Appl
29	2583.7	70.4	4660	15	US-10-464-368-74	Sequence 74, Appl
30	2580.9	70.3	4545	15	US-10-464-368-67	Sequence 67, Appl
31	2580.9	70.3	4545	15	US-10-464-368-71	Sequence 71, Appl
32	2579.9	70.3	5985	15	US-10-296-734-810	Sequence 810, App
33	2579.7	70.3	5376	14	US-10-028-248A-74	Sequence 74, Appl
34	2579.7	70.3	5376	15	US-10-107-782-74	Sequence 74, Appl
35	2578.7	70.3	5374	14	US-10-028-248A-75	Sequence 75, Appl
36	2578.7	70.3	5374	15	US-10-107-782-75	Sequence 75, Appl
37	2577.8	70.3	4655	16	US-10-741-601-314	Sequence 314, Appl
38	2577.8	70.3	4655	17	US-10-479-875-3	Sequence 3, Appl
39	2566.1	69.9	5701	9	US-09-864-761-37319	Sequence 37319, A
40	2566.1	69.9	5701	14	US-10-029-386-32438	Sequence 32438, A
41	2564.1	69.9	4495	15	US-10-138-588-20	Sequence 20, Appl
42	2560.4	69.8	5405	9	US-09-922-217-1116	Sequence 1116, Ap
43	2560.4	69.8	5405	13	US-10-025-380-1116	Sequence 1116, Ap
44	2560.4	69.8	5405	17	US-10-723-860-1647	Sequence 1647, Ap
45	2558.6	69.7	5546	15	US-10-296-734-1210	Sequence 1210, Ap

#### ALIGNMENTS

##### RESULT 1

US-10-723-860-836  
; Sequence 836, Application US/10723860  
; Publication No. US20040253606A1  
; GENERAL INFORMATION:  
; APPLICANT: Aziz, Natasha  
; APPLICANT: Gineburg, Wendy M.  
; APPLICANT: Zlotnik, Albert  
; TITLE OF INVENTION: Methods of Diagnosis of Soft Tissue Sarcoma, Compositions &  
; FILE REFERENCE: 05882.0193.NPUS01  
; CURRENT APPLICATION NUMBER: US/10/723,860  
; CURRENT FILING DATE: 2003-11-26  
; PRIOR APPLICATION NUMBER: 60/429,739  
; PRIOR FILING DATE: 2002-11-26  
; NUMBER OF SEQ ID NOS: 8393  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 836  
; LENGTH: 577  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-723-860-836

Query Match 91.3%; Score 3350.2; DB 17; Length 577;  
Best Local Similarity 86.2%; Pred. No. 1.6e-48;  
Matches 456; Conservative 36; Mismatches 19; Indels 18; Gaps 10;

QY	1	MSISDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGALVPPAALISITKGLQYV 60
DB	52	MSITSDEVNVLVRYLQESGFSHSAFTFGIESHISQSNINGTLVPPAALISITKGLQYV 111
QY	61	EAEVSINEGTLFDGRIEISLIDAVMDVVTQQAARDKLAQOH-AAAAA--- 115
DB	112	EAEISINEDGTVPDGRPIEISLIDAVMDVVTQQAARDKLAQOH-AAAAAATAA 171
QY	116	--AATNQ-QG-S---AKNGENTANGEAHTIANNHTDM-MEVDGDIPEPSKAVVLR 166

172	ATAATTTAGVSHQPNKREATVNGENRAHSV - NNHA - KPMEIDGEVEIPSSKATVLR	229
167	GHESEVFTCANPNVSDLLVSGSGSDSTARINWLSNENSTSGPTQLVLRHCIREGGQDVPSNK	226
230	GHESEVFTCANPNVSDLLVSGSGSDSTARINWLSNENSTSGPTQLVLRHCIREGGHDVPSNK	289
227	DVTSLDNMNSEGTLTATGSDYDGFARIWTKOGNLASTLGOHKGPFPALKWNRKGNFTILSAGV	286
290	DVTSLDNMNTNGTLTATGSDYDGFARIWTEDEGNLASTLGOHKGPFPALKWNRKGNFTILSAGV	349
287	DKTTIIWDAHTGEAKQKQPPFSAPALVDVQSN - TPASGSTDWCITHVCKLGQDRPIKF	345
350	DKTTIIWDAHTGEAKQKQPPFSAPALVDVQ - NNTTPASGSTDWCITHVCKLGCDRVPKVF	408
346	QGHTEVNAIKWDPTGNLLASCDDMTLKITWSMKQDNCVHDLQAHNKEIYTIKWSPTGP	405
409	QGHTEVNAIKWDPSGMLLASCDDMTLKITWSMKQEVCIHDLQAHNKEIYTIKWSPTGPA	468
406	TNNPNANMLASGSDSTVRLWDVDRGICITHTLKHOEPPYVSVAFSPDGRYLASGFDKC	465
469	TSNPNSNMLASGSDSTVRLWDIERGVCTHTLKHOEPPYVSVAFSPDGKYLASGFDKC	528
466	VHIWNTQTGALVHSYRTGGTIFEVCVNNAAGKVGASASGSGVCVLDLRK	514
529	VHIWNTOSGNLVHSYRTGGTIFEVCVNNAAGKVGASASGSGVCVLDLRK	577

```

RESULT 2
US-10-264-049-2892
; Sequence 2892, Application US/10264049
; Publication No. US20040005579A1
; GENERAL INFORMATION:
; APPLICANT: Birse et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA133P1
; CURRENT APPLICATION NUMBER: US/10/264,049
; CURRENT FILING DATE: 2002-10-04
; PRIOR APPLICATION NUMBER: PCT/US01/18569
; PRIOR FILING DATE: 2001-06-07
; PRIOR APPLICATION NUMBER: US 60/209,467
; PRIOR FILING DATE: 2000-06-07
; NUMBER OF SEQ ID NOS: 4360
; SOFTWARE: PatentIn Ver. 3.1
; SEQ ID NO 2892
; LENGTH: 542
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (45)-
; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids
US-10-264-049-2892

```

```

RESULT 3
US-10-363-829-446
; Sequence 446, Application US/10363829
; Publication No. US20040142331A1
; GENERAL INFORMATION:
; APPLICANT: Jackson, Stuart E.; Lincoln, Stephen E.;
; APPLICANT: Altus, Christina M.; Dufour, Gerard E.;
; APPLICANT: Chalup, Michael S.; Jackson, Jennifer L.;
; APPLICANT: Jones, Anissa L.; Yu, Jimmy Y.;
; APPLICANT: Wright, Rachel J.; Gietzen, Darryl;
; APPLICANT: Liu, Tommy F.; Yap, Pierre E.;
; APPLICANT: Dahl, Christopher R.; Momiyama, Monika G.;
; APPLICANT: Bradley, Diana L.; Rohatgi, Sameer D.;
; APPLICANT: Harris, Bernard; Roseberry Lincoln, Ann M.;
; APPLICANT: Gersting, Jr., Edward H.; Peralta, Careyana H.;
; APPLICANT: David, Marie H.; Panzer, Scott R.;
; APPLICANT: Flores, Vincent Z.; Daffo, Abel;
; APPLICANT: Marwaha, Rakesh; Chen, Alice J.;
; APPLICANT: Chang, Simon C.; Au, Alan P.;
; APPLICANT: Inman, Rebekah R.
; TITLE OF INVENTION: MOLECULES FOR DISEASE DETECTION AND TREATMENT
; FILE REFERENCE: PT-1183 USN
; CURRENT APPLICATION NUMBER: US/10/363,829
; CURRENT FILING DATE: 2003-03-05
; PRIOR APPLICATION NUMBER: PCT/US01/27628
; PRIOR FILING DATE: 2001-09-05
; PRIOR APPLICATION NUMBER: US 60/229,751
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,749
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,750
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,747
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/229,748
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,583
; PRIOR FILING DATE: 2000-09-05
; PRIOR APPLICATION NUMBER: US 60/230,517
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,610
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: US 60/230,597
; PRIOR FILING DATE: 2000-09-06
; Remaining Prior Application data removed - See File Wrapper or PAI
; NUMBER OF SEQ ID NOS: 506
; SOFTWARE: PERL Program

```

```
; SEQ ID NO 446
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No: L1:021759.1.orf2:2000SEP08
US-10-363-829-446

Query Match      74.0%; Score 2716; DB 16; Length 395;
Best Local Similarity 98.5%; Pred. No. 2.1e-37;
Matches 385; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 1 MSISDEVNVLVRYLQESGFSAFTFGIESHSOSNINGALVPPAALISIIQKGLQYV 60
   |||||
Db 5 MSISDEVNVLVRYLQESGFSAFTFGIESHSOSNINGALVPPAALISIIQKGLQYV 64

QY 61 EAEVSNEDGTLDPGRPIESLSLIDAVMPDVQVQROAQYRDKLAQCHAAHAAAAAATNQ 120
   |||||
Db 65 EAEVSNEDGTLDPGRPIESLSLIDAVMPDVQVQROAQYRDKLAQCHAAHAAAAAASQ 124

QY 121 QGSAKNGENTANGEENGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVEFICANPV 180
   |||||
Db 125 QGSAKNGENTANGEENGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVEFICANPV 184

QY 181 SLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 240
   |||||
Db 185 SLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWNSEGTLL 244

QY 241 ATGSDGFARITWKGNGLASTLQGHKGIPIFALKNNKKNFILSAGVDKTTIWDHAHTGEA 300
   |||||
Db 245 ATGSDGFARITWKGNGLASTLQGHKGIPIFALKNNKKNFILSAGVDKTTIWDHAHTGEA 304

QY 301 KQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 360
   |||||
Db 305 KQFPFHSAPALVDVQSNNTFASCSTDMCIHVCKLGQDRPIKTFQGHTEVNAIKWDPT 364

QY 361 GNLLASCSDDMTLKINSMKQDNCVHDLQAHN 391
   |||||
Db 365 GNLLASCSDDMTLKINSMKQDNCVHDLQAHN 395

RESULT 4
US-10-451-168-78
; Sequence 78 Application US/10451168
; Publication No. US20040091969A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; APPLICANT: SMITHKLINE BEECHAM P.L.C.
; APPLICANT: GLAXO GROUP LIMITED
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50039
; CURRENT APPLICATION NUMBER: US/10/451,168
; CURRENT FILING DATE: 2003-11-12
; PRIOR APPLICATION NUMBER: PCT/US01/49232
; PRIOR FILING DATE: 2000-12-17
; PRIOR APPLICATION NUMBER: 60/256,710
; PRIOR FILING DATE: 2000-12-19
; PRIOR APPLICATION NUMBER: 60/257,048
; PRIOR FILING DATE: 2000-12-20
; PRIOR APPLICATION NUMBER: 60/260,482
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 60/264,922
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: 60/266,797
; PRIOR FILING DATE: 2001-02-06
; PRIOR APPLICATION NUMBER: 60/276,988
; PRIOR FILING DATE: 2001-03-19
; PRIOR APPLICATION NUMBER: 60/281,535
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: 60/289,622
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 110

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 78
; LENGTH: 5635
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-10-451-168-78

Query Match      71.7%; Score 2631.9; DB 15; Length 5635;
Best Local Similarity 9.6%; Pred. No. 8.2e-30;
Matches 440; Conservative 57; Mismatches 15; Indels 4071; Gaps 372;

QY 1 M-S-----I-----SS---D-----EV----- 8
   |||
Db 1 MISWEVHTVFLFALLYSLAQDASQPSIRABEIEGASTLAFVFDVTGSMYDDLQVVI 60
   |||||
QY 9 -----NF-LV----- 12

Db 61 EGASKILETSLKRPKRPLNFALVPHDPFEPVTTTDPKKFOYELRELYVGGGDCPE 120

QY 13 -----Y-----R---Y-----LQ----- 17
   |||

Db 121 MSIGAIIKIAIEISLPGSFYVFTDASKDYRLTHEVLQIIQOKSQVVFVLTGDCDDRTH 180
   |||||
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26
   |||

Db 181 IGYKVYEIATSSGQVPHLDKQVNEVLKWEAAVQASKVHLLSTDLHLEQAVNTWRIPF 240
   |||||
QY 27 -----T-----I-----EG-----E----- 31
   |||

Db 241 DPSLKEVTVSLSGSPSMIEIRNPLGLIKKFGHLHELLNIHNSAKVNVNKEPEAGMWTVK 300

QY 32 ---S--H-----I-----S---Q----- 36
   |||

Db 301 TSSSGHSHVRIITGLSTIDFRAGFSRKPTLDFKTVSRPVQGIPTVYLLNTSGISTPARID 360
   |||||
QY 37 -----S-----NI-----N-----G----- 41
   |||

Db 361 LLELLISGSSSLKTIPIVKYYPHRKPYGIWNISDFVPPNEAFFLKVGTGYDKDYLFORVSS 420
   |||||
QY 42 -----A-----L---VP-----P-----A 47
   |||

Db 421 VFSSIVDPAPKVTMPEKTPGYVLPQGPVDSLLPFTLSFVRNGVTGLVDQYLKESA 480
   |||||
QY 48 -----A---L-----I-----S---IIQ-----KG----- 56
   |||

Db 481 SVNLDIAKVTLSDEGFYECIAVSSAGTGRAQTFDVSPPVIOVNNVTVTGERAVLT 540
   |||||
QY 57 -----L--Q-----Y---V--E----- 61
   |||

Db 541 CLIIISAVDYNLTWQRNDRVRLAEPARIRTLANLSLELKSVPKFNDAAGEYHCVSSEGGSS 600
   |||||
QY 62 -A-----EVS-----NE----- 68
   |||

Db 601 AASVFLTVQEPKVTVMKXQSGFTGSEVSIMCSATGYPKPKIAWTNDMFIVGSHRYRM 660
   |||||
QY 69 --DGTLF-----D-----G-----R---P-----I-- 78
   |||||
Db 661 TSDGTLFIKNAAPKADAGIYGCCLASNSAGTDKQNSTLRYIEAPKLMVVQSELLVALGDITV 720
   |||||
QY 79 -E---S-----L--S--LI-----DA---V----- 87
   |||

Db 721 MECKTSGIPPPQVWFKGDLRLPSTFIIDPLGLLKIQETQDLQDAGDYTCVAINBAGR 780
   |||||
QY 88 -----M-----P-DV-----VQ-----T----- 94
   |||

Db 781 ATGKITLDVGSPPVFTQEPADVSMIEGNSVTLPCYVQGYPEPTIKWRRLDNMPFISRPFS 840
   |||||
QY 95 -----R----- 95

Db 841 VSSISQIRTGALFILNLWASDKGTICEAENQFGKIQSETTIVTGLVAPLIGISPSVAN 900
   |||||
QY 96 ---QQ-----A-----Y---R-D-KL-----AQ----- 105
   |||
```

Db 901 VIEGQQLTLPCTLLAGNPERRWIKNSAMLLQNPYITVRS DGS LHI ERVQLQDGG EYTC 960  
Qy 106 -----QH-----A----- 108  
Db 961 VASVAGTNNKTTSVVHVLTPTQHQQIILSTIEGIFVTLPCASGNPKPSVMSKKGEL 1020  
Qy 109 -----A--A--A-----A--A--A----- 114  
Db 1021 ISTSSAKFSAGDGLVVSFGGESEYVCTATNAGYAKRKVOLTIVYRPRVFGDQRG 1080  
Qy 115 -----A-----A----- 116  
Db 1081 LSQDKPVEISVLAGEEVLTPCEVKS LPPITWAKETQLISPFSPRHFTLPSPGSMKITET 1140  
Qy 117 -----ATN-----Q-----Q-GSA--K-----N----- 126  
Db 1141 RTSDSGMYLCVATNIAGNVQAVKLVHVPPKIQRGPKHLKVQVGQRPDI PCNAQGTPLP 1200  
Qy 127 -----GE--N-----T--A--N--G----- 133  
Db 1201 VITWSKGGSTWLDGEHHVSNPDGTLSDQATPSDAGIYTCVATNAGTDETEILHVQE 1260  
Qy 134 -----E--N-----GA-----H----- 139  
Db 1261 PPTVEDLEPPYNTTFQERVANQRIEPPCPAKGTPKPTIKWLHNGRELTGREPGISILEDG 1320  
Qy 140 T--IA-----N-----N--HT--DM----- 148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPPVKDEQVTVNSVLLNQLTNL 1380  
Qy 149 -----M-----E----- 150  
Db 1381 FCEVEGTPSPILIMYKDNVQVTSSTIQTVNGKILKLFATPEDAGRYSCKAINAGTS 1440  
Qy 151 -----VD-----G-----DV--E-----I-----P-- 158  
Db 1441 QKYNIDVLVPTIIGTNFNEVSVLNRDVALEQVKGTPFPDIHWFKDGKPLFLGDPN 1500  
Qy 159 -----SN--K--A-----V----- 165  
Db 1501 VELLDRGQVLHLKNARNDRKRYQCTVVSNAAGQAKDIKTIYNPPSIKGNVTTDISVL 1560  
Qy 166 -----RG----- 168  
Db 1561 INSLIKLECETRGLPNPAITWYKDGQPI MSSQALYIDKQYLHI PRAQVSDSATYTCV 1620  
Qy 169 -----E-S-----E-----VFI-----C--A-----W----- 177  
Db 1621 ANVAGTAEKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECKAAGNPSPILT W LKDG V 1680  
Qy 178 -----N----- 182  
Db 1681 PVKANDNIRIEAGKKLEINSAQEI DRGQVVCVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740  
Qy 183 -----LL-----VSGS-----G--D-----S-- 191  
Db 1741 TSYFIVMNNLELDCHTVSPPTIMWLKDGQLIDERGFKILLNGRKLVI AQAVSNT 1800  
Qy 192 -----TA-----R----- 196  
Db 1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSSGLSERVVVVKYKPVALQCIANGIPNPSITW 1860  
Qy 197 -----NL--S-----E-----NS--T-----S 204  
Db 1861 LKDDQPVNTAQGNLKTQSSGRVLIQAKTLELLEDAGRYTCVATNAAGETQQHIQLHVHPPS 1920  
Qy 205 -----G-----PTQL-----V-----L-R----- 212  
Db 1921 LEDAGKMLNETLVSNPVQLECKAAGNPVPI TWYKDNRLLSGTSMTFLNRQIIDIES 1980  
Qy 213 -----H----- 215  
Db 1981 AQISDAGIYKCVAINSAGATELFLYSLQHVAPISGNSNMVAVVANNPVELECEARGIPA 2040

Qy 216 -----R-----E--G-----G--Q--D-- 221  
Db 2041 PSUTWLKDGSPVSSFNGLOVLSGGRILALTSAQISDTGRYTCVAVNAAGEKORDIDL RV 2100  
Qy 222 -VPSN-----KD-----VT--S-L- 231  
Db 2101 YVPPNMGBEQNVSVLISQAVELLQSDAIPPTLTWLKDGHPLLKKPGLSISSENSVLK 2160  
Qy 232 -----D-----W--N--S-----EG--TLL--A 241  
Db 2161 IEDAQVQDGRYTCVATNAGTTERKYNVNIWVPPNIGSGDELDTQLTVEGNLSILCES 2220  
Qy 242 T-----GS--Y-D-----G-----F-----A-- 249  
Db 2221 SGIPPNLIWKKGSPVLTDSMGRVILSGRQLOISIAEKSDAALYSCVASNAVGTAKK 2280  
Qy 250 -----R-----I-----I-----WTQD----- 256  
Db 2281 EYNLOVYIRPTITNSGSHPTETIIVTRGKSI SLECEVOGIPPTVTVWMDGHP LLI KAGVE 2340  
Qy 257 -----NL-AS-----T--L-----GOH----- 265  
Db 2341 ILDEGHILQLKNIHVS DTRGYVCVAVNAGMTDKYDLSVHAPPSIIGNHRSPENISVVE 2400  
Qy 266 K-----G--P--I--FALK--W----- 274  
Db 2401 KNSVLTCSASGIP LPSITWPF--KDGWPSVLSNSVRILSGRMLRLMQTTMEDAGQYTCV 2458  
Qy 275 -----N-----K--K-----GN----- 279  
Db 2459 VRNAAAGEERKIFGLSVLVPPIHVGENTLEDVKVKEKQSVTLTCEVTGNPVPETIWHKDGQ 2518  
Qy 280 -----P--I-----L-----S--AGV-- 286  
Db 2519 PLOEDEAHHIISGRFLQITNVQVPHTRGYTCLASSPAGKRSFSLNFVSP TIA GVS 2578  
Qy 287 D--K--T-----T--I-----I--W----- 293  
Db 2579 DGNPVDVTILNSPTSLVCEAYSPPATITWFKDGTPLESNRINIRILPGGRTLOILNAQE 2638  
Qy 294 D-----A-----H-----TG-----EA 300  
Db 2639 DNAGRYSCVATNEAGEMIKHYEVKVI PPIINKDGLWGP LSPKVKIKVNNLTLECEA 2698  
Qy 301 -----K-----Q--Q-----F 304  
Db 2699 YAI PSASLSWYKDGQPLKSDHVNIAANGHTLQIKEAQISDTGRYTCVASNIAGEDELDF 2758  
Qy 305 -----P--P----- 307  
Db 2759 DVNIQVPPSFQKLWEIGNMLDTCRNGEAKDVIINNPI SLYCETNAAPPPTLTWYKDGHP L 2818  
Qy 308 -SA-----P-----A-----LDV-----D- 315  
Db 2819 TSSDKVLILPGRGVLIQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPP I I K GANS DL 2878  
Qy 316 -----WQ-----SN----- 319  
Db 2879 PBEVTVLVNKSALIECLSSGSPAPRNSWQDQOPLLEDHDKFLSNGRILQILNTQITDI 2938  
Qy 320 -----NT-----PAS--C-----S- 326  
Db 2939 GRVYCVAEANTAGS AKKYNLVHVPPSVIGPKSENITVVVNNFISLTCEVSGFP PDL SW 2998  
Qy 327 -----T-----D--MCI----- 331  
Db 2999 LKNEQPIKLTNTLIVPGGRTLIQIRAKVSDGGEYTCIAINQAGESKKKFS LTVVYVPSI 3058  
Qy 332 --H-----V--C-----K-----L--GO-- 338  
Db 3059 KOHDSLSLVNVNREGTSVLSCEASNAPPPVITWYKNGRMITESTHVEILLADQOMLHIK 3118

QY 339 ---D---R---P-IK---TF--- 345  
Db 3119 KAEVSDTGQYVCRAINVAGRDDKNFHLNVVPPSIEGPEREVIVETISNPVTLTCDATGI 3178  
QY 346 ---Q---G---H---T--- 349  
Db 3179 PPPTIAWLKHKRIENSLEVRLLSGGSKLOIARSOHSDSGNYTCIASNMEGKAQYYF 3238  
QY 350 ---N-E-V-NA---IKW---D-P---T--- 360  
Db 3239 LSIQVPSVAGAEIPSDVSVLLGENVELVCNANGIPTPLQWLKDGKPIASGETERIRVS 3298  
QY 361 ---G---NLL-A-S---C-S---D---D---MTL--- 373  
Db 3299 ANGSTLNIYCALSDTGKYTCVATNPAGEEDRIFNLNVVPTPIRGNKDEAKLMTLVDI 3358  
QY 374 ---K---I-W---SM--- 378  
Db 3359 SINIECRATGPPQINWLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVVTCVASNR 3418  
QY 379 ---K---Q---DN---CV--- 384  
Db 3419 AGVDNKHYNLOVPAPPNMDNSMGTEETVLKGSSTWACITDGTAPAPMAWLEDGQPLGL 3478  
QY 385 ---H---D---L---QA---H-N---KEI 394  
Db 3479 DAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKHFLKLVLEPPHINGSBEEHEI 3538  
QY 395 ---YT---I---K-W--- 399  
Db 3539 SVIVNNPLEITCIASGAPKMTWMKGRPLPOTQVOTLGGGEVLRISTAQVEDTGRYT 3598  
QY 400 ---SPTG---P---GT---N--- 408  
Db 3599 CLASSPAGDDDKBYLVRVHPVPIAGDEPRDITVLNRQVLTCEKSDAVPPPVIITWLRN 3658  
QY 409 ---P---N---ANLM-LAS--A---S-F--- 420  
Db 3659 GERLOATPRVRLISGGRYLQINNADLGDNTANYTCVASNIAGKTRREFILTVNVPPIKGG 3718  
QY 421 ---DSTV---RL-W-D-V---DR---G-I--- 433  
Db 3719 POSLVLINKSTVLECIAEGVPTPRITWRKOGAVLAGNHARYSILENGFLHIOSAHVTDI 3778  
QY 434 ---C---I---H---CV---TL---T--- 439  
Db 3779 GRYLCMATNAGTDRRIDLOVHPVPSIAGPNTMTVINVQTLTACEATGIPKPSINWR 3838  
QY 440 K-H---Q-E---P-V---Y---S 447  
Db 3839 KNGHLLNVQDNQNSYRLSSGSLVLIISPSYDDTATYECTVNGAGDDKRTVDLTVOVPPS 3898  
QY 448 VA---F-S---P-D--- 453  
Db 3899 IADEPTDFLVTHAPAVITCTASGVPPSPSIHWTNGIRLLPRGDGYRLLSSGAIEILATQ 3958  
QY 454 ---GRY---L---SG--S-F 462  
Db 3959 LNHAGRYTCVARNAGSAHRVTLHVHPEPVIQPSSELVILNPIILLPCEATGTPSPF 4018  
QY 463 ---D---KCV--- 466  
Db 4019 ITWQKEGINVNTSGRNVHVLPSGGLQISRAVREDAGTYMCVAQNPGATLKGKILNVQVP 4078  
QY 467 ---H---I-W--- 469  
Db 4079 PVISPHLKEVIAVDKPIITLSCADGLPPDDITWHKDGRAIVESIRQVLSGSSGLQIAFV 4138  
QY 470 ---N---T---Q--- 472  
Db 4139 QPGDAGHYTCMAANVAGSSSTSTKLTVHPVPRIRSTEGHYTVNNSQAILPCVADGIPTP 4198  
QY 473 ---T---G---A---L-VH- 478

Db 4199 AINWKDNVLLANLLGKYTAEPYGELILENVVLEDSGFYTCVANNAAGEDTHTVSLTVHV 4258  
QY 479 ---S---Y---R- 481  
Db 4259 LPTFTLPGDVSINKGEQLRLSCKATGIPLPKLTWTFNNNIIIPAFHDSVNGHSELVIERV 4318  
QY 482 ---GT---G---G---IF---E---V-C--- 490  
Db 4319 SKEDSSTYVCTAENSUGFVKAIGFVVVYVVPFKGDPSPNWIPLGNGAILNCEVKGDP 4378  
QY 491 ---WN---A---AGD--- 496  
Db 4379 PTIQNRKGVDIISHIRIQLNGSLAIYGTVNEADAGDYTCVATNEAGVVERSMSTLQS 4438  
QY 497 ---K-V---G---A---SAS--- 503  
Db 4439 PPIITLPEVETINAGKILNCOATGEPQPTITWSRQSHISWDDRVNVLSNNSLYIAD 4498  
QY 504 ---D---GSV---C-V--- 509  
Db 4499 AQKEDTSEFECVARNLMSGVLVRVPVIVQVHGGFSQMSAWRACSVTCGKGIQKRSLCNQ 4558  
QY 510 -L-----DL--R-----K 514  
Db 4559 PLPANGGKPCQSGDLEMRNCQNK 4581

## RESULT 5

US-10-032-189-128  
; Sequence 128, Application US/10032189  
; Publication No. US20030170630A1  
; GENERAL INFORMATION:  
; APPLICANT: Alsobrook II, John P  
; APPLICANT: Tchernev, Velizar T  
; APPLICANT: Liu, Xiaohong  
; APPLICANT: Spytek, Kimberly A  
; APPLICANT: Zerhusen, Bryan D  
; APPLICANT: Patturajan, Meera  
; APPLICANT: Grosse, William M  
; APPLICANT: Lepley, Denise M  
; APPLICANT: Burgess, Catherine E  
; APPLICANT: Shimkets, Richard A  
; APPLICANT: Grosse, William M  
; APPLICANT: Szekeres, Edward S  
; APPLICANT: Vernet, Corine A.M.  
; APPLICANT: Li, Li  
; APPLICANT: Casman, Stacie J  
; APPLICANT: Boldog, Ferenc L  
; APPLICANT: Gorman, Linda  
; APPLICANT: Gangolli, Esha A  
; APPLICANT: Fernandes, Elma R  
; APPLICANT: Rieger, Daniel K  
; APPLICANT: Edinger, Shlomit R  
; APPLICANT: Gunther, Erik  
; APPLICANT: Millet, Isabelle  
; APPLICANT: Sciore, Paul  
; APPLICANT: Ellerman, Karen  
; APPLICANT: MacDougall, John R  
; APPLICANT: Smithson, Glenna  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-228  
; CURRENT APPLICATION NUMBER: US/10/032,189  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/257,495  
; PRIOR FILING DATE: 2000-12-21  
; PRIOR APPLICATION NUMBER: 60/258,171  
; PRIOR FILING DATE: 2000-12-20  
; PRIOR APPLICATION NUMBER: 60/269,940  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: 60/274,192  
; PRIOR FILING DATE: 2001-03-08  
; PRIOR APPLICATION NUMBER: 60/277,826

; PRIOR FILING DATE: 2001-03-22  
 ; PRIOR APPLICATION NUMBER: 60/279,840  
 ; PRIOR FILING DATE: 2001-03-29  
 ; PRIOR APPLICATION NUMBER: 60/282,981  
 ; PRIOR FILING DATE: 2001-04-11  
 ; PRIOR APPLICATION NUMBER: 60/283,656  
 ; PRIOR FILING DATE: 2001-04-13  
 ; PRIOR APPLICATION NUMBER: 60/309,247  
 ; PRIOR FILING DATE: 2001-07-31  
 ; PRIOR APPLICATION NUMBER: 60/311,754  
 ; PRIOR FILING DATE: 2001-08-17  
 ; PRIOR APPLICATION NUMBER: 60/313,331  
 ; PRIOR FILING DATE: 2001-08-17  
 ; NUMBER OF SEQ ID NOS: 260  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 128  
 ; LENGTH: 5636  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: (3003)  
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: (3041)  
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.  
 ; FEATURE:  
 ; NAME/KEY: VARIANT  
 ; LOCATION: (3367)  
 ; OTHER INFORMATION: Wherein Xaa is any amino acid.  
 ; US-10-032-189-128

Query Match 71.6%; Score 2627.8; DB 14; Length 5636;  
 Best Local Similarity 9.6%; Pred. No. 9.8e-30;  
 Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;  
 1 M-S-----Y-----SS--D-----EV-----8  
 1 MISVEVHTVFLFALLYSSLAQASQSEIRAEFFEGASTLAFVDPVTGSMYDDLQVVI 60  
 9 -----NF-LV-----12  
 61 EGAKILETLKRPKRPLNFALVPHDPHDPBPGVTTITDPKQFYLRELYVQGGDCPE 120  
 13 -----Y-----R--Y--LQ-----17  
 121 MSGAIKALEISLPGSFYVFTDARKSKOYRLTHEVLQLIQKOSQVVFVLGTGCDRTH 180  
 18 -----E-----SG-F-----S-H-S-----A-----F 26  
 181 IGVKVEEIASTSGQVFHLDKKQVNEVLKWEVQAEVQSKVHLLSTDHLEQAVNTRIPF 240  
 27 -----T-----FG-----I-----E-----31  
 241 DPSLKEVTVSLGSPSMIEIRNPLGLIKKGFGHLHELLNIHNSAKVNVVKEPEAGWTVK 300  
 32 ---S-H-----I-----S-Q-----36  
 301 TSSSGRHSVRITGLSTIDFRAGFSRKPFLDPKTVSRPVQGIPTYVLLNTSGISTPARID 360  
 37 -----S-----NI-----N-----G-----41  
 361 LLELLSISGSSSLKTI PVKYYPHRRKPYGIWNI SDVPNPNEAPFLKVTGYDKDYLFQVSS 420  
 42 -----A-----L-----VP-----P-----A 47  
 421 VSFSSIVPDAPKVTMPKTPGYVLQPGQIPCSVDSLPLPFTLSFVRNGVTLGVQYLKESA 480  
 48 -----A--L-----I-----S-----IIO-----KG-----56  
 481 SVSLDIKAVTLSDGFEYECIAVSAGTGRAQTFFDVSEPPPIQVPPNNVTVPGERAVLT 540

QY 57 -----L-Q-----Y---V-E---61  
 DB 541 CLIIISAVDYNLTWQRNDRVRLAEPARIRTLANLSLELKSVKFNDAGEYHCVSSEGSS 600  
 QY 62 -A-----EVS-I-----NE-----68  
 DB 601 AASVFLTVQEPKPKVTMPKNQSF TGGSEVSI MCSATGYPKPKIAMTVNDMFIVGSHRYRM 660  
 QY 69 --DGLTF-----D-----G-----R--P-----I--78  
 DB 661 TSDGTLFIKNAAPKADAGIYGCILASNSAGTDKQNSTURYEAPKLMVQSELLVALGDITV 720  
 QY 79 -E---S-----L-S-LI-----DA-----V-----87  
 DB 721 MECKTSGIPPPQVWKFGDLELRPSTFLIIDPLGLLKLKIQTQDLADAGYTCVAINAAGR 780  
 QY 88 -----M-----P-DV-----VQ-----T-----94  
 DB 781 ATGKITLDVGGSPPVFIQEPADVSMIEIGSNVTLPCYVQGYPEPTIKWRRLDNMFI SRPFS 840  
 QY 95 -----R-----95  
 DB 841 VSSISQLRTGALFILNLWASDKGTIYCEABNOFGKIQSETTVTGTGLVAPLIGISPVAN 900  
 QY 96 ---QQ-----A-----Y---R-D-KL---AQ-----105  
 DB 901 VIEGQQLTLPCTLLAGNPERRWIKNSAMLLQNPHYITVRS DGLHIERVQLQDGGEYTC 960  
 QY 106 -----OH-----A-----108  
 DB 961 VASNVAGTNKTTVVVHVLTQIHHQQLSTIEGIPVTLPCCKASGNPKPSVIWKKGEL 1020  
 QY 109 ---A---A-A-----A-----A-----114  
 DB 1021 ISTSSAKFAGADGSLYVVSPEGESGEYVCTATNTAGYAKRKVQLTVVYRPRVFGDLRG 1080  
 QY 115 -----A-----A-----116  
 DB 1081 LSQDKPVEISVLAGEBVTLPCEVKSLPPPIITWAKETQLISPFSPRHTLPSPGSMKITET 1140  
 QY 117 -----NTN-----Q-----O-GSA-K-----N-----126  
 DB 1141 RTSDSGMYLCVATNTAGNTQAVKLVNHPVKIQRGPKHLKVQGVORVDPICNAQGTPLP 1200  
 QY 127 -----GE--N-----T-A-N-G-----133  
 DB 1201 VITWSKGGSTMLVDGEHHVSNPDGTLSDIQA TPDSAGIYTCVATNIAGTDETEILHVQE 1260  
 QY 134 -----E--N-----GA-----H-----139  
 DB 1261 PPTVEDLEPPYNTTFOERVANQRIEPPCPAKGTPKPTIKWLHNGRELTGREPGISILEEG 1320  
 QY 140 T--IA-----N-----N--HT---DM-----148  
 DB 1321 TLLVIATSPYDNGEYICVAVNEAGTTERKYNLKVHVPVPIKDKEQVSNVLLNQLTNL 1380  
 QY 149 -----M-----E-----150  
 DB 1381 FCEVEGTPSPIIMWYKDNQVVTESSTIQVNNKGILKLFPRATPEDAGRYSCKAINAGTS 1440  
 QY 151 -----VD-----G-----DV-E-----I-----P-158  
 DB 1441 OKYFNIDVLVPPTIIGTNFPEKVS VVNLNRDVALEQCVKGTPPDIHWFKDGKPLFLGDPN 1500  
 QY 159 -----K-A-----SN-----K-A-----V-----VL 165  
 DB 1501 VELLDRGQVLHLKNARRNDKGRYQCTVSNAAAGQAKDIKLTIIYIPPSIKGNVTDTISVL 1560  
 QY 166 -----KG-----H-----168  
 DB 1561 INSILKLECKTRGLPMPALTWYKDGQPI MSSSQALYIDKGQYLHIIPRAQVSDSATYCHV 1620  
 QY 169 -----E-S-----E-----VFI-----C-A-----W-----177



Db 1621 ANVAGTAESFHDVYVPPMIEGNLATPLNKQVVIASHLTLECAAGNPSPIITLWKDGV 1680  
QY 178 -N- - - - -P-V-SD- 182  
Db 1681 PVKANDFRIEAGKKLEINSAQEIQRGGYICVATSVAGEKEIYEVVDVLVPPAIBGGDE 1740  
QY 183 -LL- -VSGS- -G-D- -S- 191  
Db 1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKDGQILDERDGFILLNGRKLIVIAQAQVNT 1800  
QY 192 -TA- - - - -R- - - - -I-W 196  
Db 1801 GLYRCMAANTAGHKKEFEVTVHPPTIKSSGLSERVVVVKYKPAVLCQIANGIPNSITW 1860  
QY 197 -NL- -S- -E- -NS- -T- -S 204  
Db 1861 LKDDQPVNTAQNKKQSSGRVLIQAKTLEDAGRYTCVATNAAGETQOHIQLHVHPPS 1920  
QY 205 -G- -PTQL- -V- -L-R- - 212  
Db 1921 LEDAGKMLNETVLSNPVQLECKAAGNPVPIITWYKDNCLLSGTSMTFLNRQIIDIES 1980  
QY 213 -H- - - - -C- -I- - 215  
Db 1981 AQISDAGIYKCAINSAGATELFYSQVHVAPISGSNNMVAVVVNNPVRLECEARGIPA 2040  
QY 216 -R- - - - -E- -G- -G-Q- -D- - 221  
Db 2041 PSITLWKDGPVSPFNSGLQVLSGGRILALTSTQISDTRYTCVAVNAAGEKORDILRV 2100  
QY 222 -VPSN- - - - -KD- - -VT- -S-L- 231  
Db 2101 YVPNMGEBEQNVLSIQAVELLQSDAIPPTTLTWLKDGHLLKPKGLSISENSVLK 2160  
QY 232 -D- - - - -W- -N- -S- -EG- -TLL- 240  
Db 2161 IEDAQVQDTGRYTCEATNVAGTEKKYNNIWWPNIGSGDELQTLVIEGNLILCE 2220  
QY 241 AT- - - - -GS- -Y-D- -G- -F- - -A- 249  
Db 2221 SSGIPPNLWKKGSPVLTDSMGRVRLSGGRQLQISIAEKSDAALYSCVASNAGTAK 2280  
QY 250 -R- - - - -I- - -WTKDG- - 256  
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSISECEVOGIPPTVTWTKDGHLLKAKGV 2340  
QY 257 -NL-AS- - - - -T- -L- -GOH- - 265  
Db 2341 EILDEGHILQKNIHVSDTRYVCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400  
QY 266 -K- - - - -G-PI- - -FALK- -W- - - - 274  
Db 2401 EKNSVLTCEASGIPLPSTTFW- -KOGWPVLSNSVRILSGGRMLRMQTTMEDAQGYTC 2458  
QY 275 -N- - - - -K-K- - -GN- - - 279  
Db 2459 VVNAAGEERKIFGLSVLPVPHVIGENTLEDVKVKEQSVTLTCEVTGNPVPITWHKDG 2518  
QY 280 -F-I- - - - -L- -S- -AGV- 286  
Db 2519 QPLQDEAHIIISGRFLQITNVQVPHGTGRYTCLASSPAGHKRSFSLNVFVSPTIAGVG 2578  
QY 287 -D- -K-T- - - - -I-W- - - 293  
Db 2579 SGNPEDVTILNSPTSLVCEAYSYPATITWFKDGTPLSNRNIRILPGGRTLOILNAQ 2638  
QY 294 -D- - - - -A- -H- - -TG- - -E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWGLGSPKEVKIKVNTLTLECE 2698  
QY 300 A- - - - -K- - - - -Q- - - - 303

Db 2699 AVAIPASLSWYKDGQPLKSDDDHVNIAANGHTLIQKEAQISDTRYTCVASNIAGEBELD 2758  
QY 304 F- - - - -P-F- - - - -H- 307  
Db 2759 FDNIOVPPSFQKLWEIGNMLDTGRNGEAKDVIINNPIISLYCETNAAPPTLTWYKDGHP 2818  
QY 308 -SA- - - - -P- - - - -LDV- - -D 315  
Db 2819 LTTSSDKVLILPGGRVLIQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPPIEGANS 2878  
QY 316 -WQ- - - - - -SN- - - - - 319  
Db 2879 LPPEVTVLNKSALIECLSSGSPAPRNSWQKQPLLEDDHHKFLSNGRILQILNTQITD 2938  
QY 320 -NT- - - - -FAS- -C- -S 326  
Db 2939 IGRYCVCAENTAGSAKKYFNLVNHHVPPSVIGPKSENLTVVNNFISLTCEVSGFPDPLS 2998  
QY 327 -T- - - - -D- - -MCI- - - 331  
Db 2999 WLKNXQPIKLTNTLIVPGGRTLQIIRAKVSDGGEYTCIAINKAGESKKFSLTVVPPS 3058  
QY 332 -H- - - - -V-C- -K- -L-GQ- - 338  
Db 3059 IKDHDSESLVVNREGTSVLECESNAVPPVITWYKNGRMITESTHVEILLADGQMLHI 3118  
QY 339 -D- - - - -R- - - -P-IK- - - 343  
Db 3119 KKAESDTCQYVCRAINVAGRDDKNFNLNVPVPSIEGPEREVIVETISNPVTLTCDATG 3178  
QY 344 -T- - - - -F- - - -Q-G- -H- - - 349  
Db 3179 IPPPTIAMLKNYKRIENSLSLEVRILSGSKLQIARSQSDSGNYTCIASNMEGKAKYY 3238  
QY 350 -N-B-V-NA- - - - -IKW- -D- -P- - -T- - 360  
Db 3239 FLISQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTLQIWLKDGKFIASGETERIRV 3298  
QY 361 -G- -NL-A- -S- - - -C- -S- - -D- - - -D- - -MTL- - 373  
Db 3299 SANGSTLNIYALTSDTGKYTCVATNPAGEEDRIFNLVNVPVITIRGNKDEAKLMTYVD 3358  
QY 374 -K- - - - -I-W- - - -SM- - - - 378  
Db 3359 TSINIECRXTGTPPQIINLKNGLPLPLSSHRLLAAGOVIRIVRAQVSDVAVVTCVASN 3418  
QY 379 -K- - - - -Q- - - -DN- - - -CV- - - - 384  
Db 3419 RAGVDNKHYNLQVFAFPNNDMSMGTEETVLKSGSSTSMACITDGTAPSMAWLRDQPLG 3478  
QY 385 -H- - - - -D- - - -L- - - -QA- - - -H-N- - - -KE 393  
Db 3479 LDAHLTVSTHGMVLQLLKAETSDSGKYTCIASNEAGEVSKHFLKVLPEPHINGSEHEE 3538  
QY 394 I- - - - -YT- - - - -I-K- -W- - - -N- - - - 399  
Db 3539 ISVIVNNPLELTIASGIPAPKMTWMDKGRPLPQTDQVTLGGGEVLRISTAQVEDTRY 3598  
QY 400 -SPTG- - - - -P- - - -GT- - - - - 407  
Db 3599 TCLASSPAGDDKEYLVRVHVPPNIAGTDEPRDITVLRNRQVLTLECKSDAVPPPVTWLR 3658  
QY 408 N- - - - -P- - - - -N- - - -ANLM-LAS- -A- -S- -F- - - - 420  
Db 3659 NGERLOATPRVILSGRYLQINNADLGTANTYTCVASNIAGKTTREFILTNNVPPNKG 3718  
QY 421 -DSTV- - - - -RL-W- -D- -V- - - -DR- - - -G- -I- - - 433  
Db 3719 GPQSLVILLNKSTVLECIAGVPTPRIWTKDQAVLGNHARYSILENGLHQAHAHTD 3778  
QY 434 -C- - - - -I- - - -H- - - -TL- - - -T- - - 439  
Db 3779 TGRYLCMATNAAGTDRRRIDLQVHPPSIAPGTNNMTVNVVQTTLACEATGIPKPSINW 3838



QY 48 A--L--I--S--IIQ--KG-- 56  
Db 481 SVSLDIKAVTLSDGFEYCIASVAGTGRAQTFDVSPPVPIQVNNVTVTPGERAVLT 540  
QY 57 L--Q--Y--V--E-- 61  
Db 541 CLIIASVDYNTLWQRNDRVRLAEPARTLANLSLELKSVPKNDAGEYCHVMSSEGGSS 600  
QY 62 A--EVS1--NE-- 68  
Db 601 AASVFLTVQEPKVTWPKMQSTGSEVSIMCAGYKPKIAWTVMDFIVGSHRYM 660  
QY 69 DGTLP--G--R--P-- 78  
Db 661 TSDGTLFKNAAPKDIAGIYGLASAGTDKQNSTLRYEAPKLMVVOSELLVALGDITV 720  
QY 79 E--S--L--S--LI--DA--V-- 87  
Db 721 MECKTSGIPPOVKWFKGDLRLPSTFLIIDPLGLLKIQETQDLADGDTCTVAINAAGR 780  
QY 88 M--P--DV--T--VO-- 94  
Db 781 ATCKITLDVCSPPVFIQEPADVMEIGSNVTLCYVQGYPEPTIKWRRLDNMPIRSPPS 840  
QY 95 R-- 95  
Db 841 VSSISQRLTGALFILNLWASDKGTIYCEAENQFGKIQSETTVTVTGLVAPLIGISPSVAN 900  
QY 96 QO--A--Y--R--D--KL--AQ-- 105  
Db 901 VIEGQOULTPCTLLAGNPIERRWIKNSAMLLQNPYITVRSGLSHIERVOLODGGEYTC 960  
QY 106 OH--A-- 108  
Db 961 VASNAVAGTNKTSVVVHVLPITIQHQOILSTTEGIPVTLPCASGNPKPSVWSKKGEL 1020  
QY 109 A--A--A--A--A-- 114  
Db 1021 ISTSSAKFSAGADSLVVSPEGESGEYCTATNTAGYAKRVQLTVVYRPFVGLRG 1080  
QY 115 A--A-- 116  
Db 1081 LSQDKPVEISVLAGEEVLTPCEVKSUPPIITWAKETQLISPPSPRHTFLPSPGSMKITET 1140  
QY 117 ATN--Q--Q--GSA--K--N-- 126  
Db 1141 RTSDSGMYLCVATNIAGNVTQAVKLVNHVPPKIQRGPKHLKVQVQORVDIPCAQGTPLP 1200  
QY 127 GE--N--T--A--N--G-- 133  
Db 1201 VITWSKGSNMLVDGHHVSNPDGTLSDOATPSDAGIYTCVATNIAGTDETEILHVQE 1260  
QY 134 E--E--N--GA--H-- 139  
Db 1261 PPTVEDLEPPYNTTFQERVANORIEPCPAKTPKPTIKWLHNGRELTGREGSILEEG 1320  
QY 140 T--IA--N--N--HT--DM-- 148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKDKQVSNVSVLLNQLTNL 1380  
QY 149 M--E-- 150  
Db 1381 FCEVEGTPSPFIWYKDNVQVTESSITQTVNNGKILKLFRATPEDAGRYSCKAINAGTIS 1440  
QY 151 VD--G--DV--E--I--P-- 158  
Db 1441 QKYFNIDVLPPTIIGTFPKEVSVNLNRDVALEQCVKGPFPFDIIHFKDGKPLFLGDPN 1500  
QY 159 SN--K--A--V--VL 165  
Db 1501 VELLDRGOVLHLKNARNRNDKGRYQCTVSNAAQKADIKLTIYIPPSIKGNVTTDISVL 1560

QY 166 RG--H-- 168  
Db 1561 INSILKLECKTRGLPMPAITWYKDGQPIIMSSSQALYIDKQYLHIIPRAQVSDSATYTCV 1620  
QY 169 E--S--E--VFI--C--A--W-- 177  
Db 1621 ANVAGTAESFHVVDVVPPIEGNLATPLNKQVVIASHLTLECNAAAGNPSPILTWIKDGV 1680  
QY 178 N--P--V--SD-- 182  
Db 1681 PVKANDNFRIEAGGKLEIMSAQIDRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740  
QY 183 LL--VSGS--G--D--S-- 191  
Db 1741 TSYFIVMNNLLELDCHVTGSPPTIMWLKQGLIDERDGFKILLNGRKLVIATAQAQVSNT 1800  
QY 192 TA--R--I--W 196  
Db 1801 GLYRCMAANTAGDHKEFEVTVHPPTIKSSGLSERVVVKYKPVALQCIANGIPNSITW 1860  
QY 197 NL--S--E--NS--T--S 204  
Db 1861 LKDDQPVNTAQGNLKIQSSGRVLIQIAKTILLEDAGRYTCVATNAAGETQQHIQLHVHPPS 1920  
QY 205 G--PTOL--V--L--R-- 212  
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPVITWYKDNCLLSGSTSMFTFLNRQIIDIES 1980  
QY 213 H--C--I-- 215  
Db 1981 AQISDAGIYKCVAINSAGATELFYSLQHVAPISGSNNMVAVVNNPVRLCEARGIPA 2040  
QY 216 R--G--E--G--D-- 221  
Db 2041 PSITWLKDGSPVSSFNGLQVLSGGRILALTSTQISDTGRYTCVAVNAAGEKORDILRV 2100  
QY 222 VPSN--KD--VT--S--L-- 231  
Db 2101 YVPPNIMGEEQNVSVLISQAVELLQSDAIPPTLTWLKDGHPHLLKPKGLSISENSVLK 2160  
QY 232 D--W--N--S--EG--TLL-- 240  
Db 2161 IEDAQVDTGRYTCBATNVAGKTEKKNVNNVWPPNIGSGDELQTLTIEGNLISLCE 2220  
QY 241 AT--GS--Y--D--G--F--A-- 249  
Db 2221 SSGIPPPNLIWKKGSPVLTDSMGRVRIILSGRQLOISAEKSDAALYSCVASNVAGTAK 2280  
QY 250 R--I--WTXDG-- 256  
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSISLECEVQGIPTPTVTWMDGHPHLLKAKGV 2340  
QY 257 NL--AS--T--L--GQH-- 265  
Db 2341 EILDEGHILQKNHIVSDTGRYVCVAVNVAGMTDKYDLSVHAPPSIIGNHRSPEISVV 2400  
QY 266 K--G--PI--FALK--W-- 274  
Db 2401 EKNSVLTCEASGIPLPSTTWP--KDGWPSVLSNSVRIILSGGRMLRLMQTTMEDAQYTC 2458  
QY 275 N--K--K--GN-- 279  
Db 2459 VVRNAAAGEBKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPBITWHKDG 2518  
QY 280 F--I--L--S--AGV-- 286  
Db 2519 QPLQEDAEHHIISGGRFLQITNVQVPHGTCTCLASSPAGKSRSPSLNVFVSPPTIAGV 2578  
QY 287 D--K--T--T--I--I--W-- 293  
Db 2579 SGNPEDVTILNSPTSLVCEAYSPPATITWFKDGTPLESNRNIRILPGRTLOILNAQ 2638  
QY 294 D--A--H--TG--E 299

Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIINKGDLWGGLSPKVEKIKVNNITILECE 2698  
QY 300 A-----K-----Q-----Q----- 303  
Db 2699 AVAIPASASLWYKDGQPLKSDHVNIAANGHTLIQKEAQISDTRVTCVASNIAGEDELD 2758  
QY 304 F-----P-----H-----H----- 307  
Db 2759 FVNIQVPPSFQKLWEIGNMLDTRNGEAKDVIINNPIISLYCETNAAPPPTLTWYKDGHP 2818  
QY 308 -SA-----P-----A-----LDV-----D 315  
Db 2819 LTSQDKVILPGRVLOIPRAKVEDAGRYTCVAVNEAGESLSQYDVRVLVPIIEGANSND 2878  
QY 316 -----WO-----SN----- 319  
Db 2879 LPBEVTVLVNKSALIECLSSGSPAPRNSWQKDGQPLEDDHKKFLSNGRILQILNTQITD 2938  
QY 320 -----NT-----PAS-----C-----S 326  
Db 2939 IGRYVCVAENTAGSAKRYFNLNVHVPSPVIGPKSENLTVVVNNFISLTCEVSGFPDPLS 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKNXQPIKLTNTLIVPGGRTLIQIRAKVSDGGEYTCIAXGSKKKFSLTVVYPPS 3058  
QY 332 -H-----V-----C-----K-----L-GQ----- 338  
Db 3059 IKHDSLSLVNVREGTSVLECSNAVPPVITWYKNGRMITESPHVEILADGQMLHI 3118  
QY 339 -D-----R-----P-IK----- 343  
Db 3119 KKAESDGTGOYVCRAINAGRDKNFHLNVYVPPSIEGPEREIVETISNPVTLTCDATG 3178  
QY 344 -T-----F-----Q-----G-----H-----T----- 349  
Db 3179 IPPPTIAWLKNYKRIENSLSLEVRILSGGKLOIARSOHSDSGNYTCIASNMEGKAQKY 3238  
QY 350 -----G-----N-E-V-NA-----IKW-D-P-----T----- 360  
Db 3239 FLSIQVPPSVAGAEIPSDSVLLGENVELVCNANGIPTPLIQWKDKGPIASGETERIRV 3298  
QY 361 -G-----NL-A-S-----C-S-----D-----D-----MTL-- 373  
Db 3299 SANGSTLNIYGALTSDTGKYTCVATNPAGEEDRIFNLNVYVTPTRGNKDEAEKLMTYD 3358  
QY 374 -K-----I-W-----SM----- 378  
Db 3359 TSINIECRXTGTPPPQINWLNKGLPLPLSHIRLLAAGVIRIVRAQVSDVAVYTCVASN 3418  
QY 379 -K-----Q-----DN-----CV----- 384  
Db 3419 RAGVDKNHYNLQVAPPNMDSNGTEBITVLKGSSTSMACITDGTAPASMAWLRDGOPLG 3478  
QY 385 -H-----D-----L-----QA-----H-N-----KE 393  
Db 3479 LDAHLTVSTHGMVLQKAEETDSGKYTCIASNEAGEVSKHFLKVLPEPHINGSSEHEE 3538  
QY 394 I-----YT-----I-K-W----- 399  
Db 3539 ISVIVNPNLELTCIASGIPAKMTWMDGRPLPQTDQVQTLGGGEVLRISTAQVEDTGRY 3598  
QY 400 -----SPTG-----P-----GT-----N----- 407  
Db 3599 TCLASSPAGDDKEYLVRHVHPNIACTDEPRDITVLNRQVLTCKSDAVPPPVITWLR 3658  
QY 408 N-----P-----N-----ANLM-LAS-A-S-F----- 420  
Db 3659 NGERLOATPRVILSGRYLOINNADLGDYANTCVASNIAGKTRETFILTVNVPNNIKG 3718  
QY 421 -----DSTV-----RL-W-D-V-----DR-----G-I----- 433

Db 3719 GPQSLVILLNKSTVLECIAEGVPTPRI TWKDGAVLGNHARYSILENGFLHIQSAHVTD 3778  
QY 434 -----C-----I-----H-----TL-----T----- 439  
Db 3779 TGRYLCWATNAAGTDRRRIDLQVHVPPSIAPGFTNMTVINNVQTTLLACEATGIPKPSINW 3838  
QY 440 -K-H-----O-E-----P-V-----Y----- 446  
Db 3839 RKNGHLLNVDOQNSYRLLSSGSLVIISPSVDDTATYEYTVNGAGDDKRTVDLTVQVPP 3898  
QY 447 SVA-----F-S-----P-D----- 453  
Db 3899 SIADPTDFLVTKHAPAVITCTASGVPPFSIHWTKNIGIRLLPRGDGYRILSSGAIBLAT 3958  
QY 454 -----GRY-----L-----ASG-S- 461  
Db 3959 QLNHAGRYTCVARNAAGSAHRHVTLHVHBPVPIQOPSELHVLINPNILLPCEATGTPSP 4018  
QY 462 F-----D-----KCV----- 466  
Db 4019 FITWQKEGINVNTSGRNHVLPSGGLQISRAVREDAGTMCVAQNAGTALGKILNVQV 4078  
QY 467 -----H-----I-W----- 469  
Db 4079 PPVISPHLKEYVIAVDKPTILSCADGLPPDITWHKDGRAIVESIRQVLSGSLQIAF 4138  
QY 470 -----N-----T-----Q----- 472  
Db 4139 VOPGDAGHYTCMAANVAGSSSTKLTVHVPPIRSTEGHYTVNENSOAILPCVADGPT 4198  
QY 473 -----T-----G-----A-----L-VH 478  
Db 4199 PAINWKKONVLLANLLGKYTABPYGELILENVVLEDSGFYTCVANNAGEDHTVSLTVH 4258  
QY 479 -----S-----Y-----R 481  
Db 4259 VLPTFTTELPGDVLNKGELRLSKATGIPKLWTNNNIIPAHFDSVNGHSELVIER 4318  
QY 482 -----GT-----G-----IF-----E-----V-C----- 490  
Db 4319 VSKEDSGTYVCTAENSVGFKATGTFYVYKPEPVFKGDPFNSNWIPLGNAILNCEVKGDP 4378  
QY 491 -----WN-----A-----AGD----- 496  
Db 4379 TPTIQNRKGVDIIEISHRIQLNGSLAIYGTWNEDAGTYTCVATNEAGVVERSMSTLQ 4438  
QY 497 -----K-V-----G-----A-----SAS----- 503  
Db 4439 SPPIITLEPVETVINAGGKIILNCQATGEPQPTITWSRQGHSSISWDDRVNLSNLSLIA 4498  
QY 504 -----D-----GSV-----C-V----- 509  
Db 4499 DAQKEDTSBECVARNLMSVLVRVPIVQVHGFSGNSAWRACSVTCGKIQKRSLCN 4558  
QY 510 -L-----DL-R-----K 514  
Db 4559 QPLPANGKPKQGSDDLEMRNCQNK 4582

## RESULT 7

US-10-023-634-93

; Sequence 93, Application US/10023634

; Publication No. US20030236389A1

; GENERAL INFORMATION:

; APPLICANT: Shimkets, Richard A

; APPLICANT: Colman, Steven D

; APPLICANT: Spoytek, Kimberly A

; APPLICANT: Ballinger, Robert A

; APPLICANT: Guo, Xiaojia

; APPLICANT: Tchernev, Velizar T

; APPLICANT: Shenoy, Suresh G

; APPLICANT: Li, Li

; APPLICANT: Ellerman, Karen

APPLICANT: Zerhusen, Bryan D  
APPLICANT: Patturajan, Meera  
APPLICANT: Casman, Stacie J  
APPLICANT: Boldog, Ferenc  
APPLICANT: Gusev, Vladimir Y  
APPLICANT: Burgess, Catherine E  
APPLICANT: Edinger, Shlomit R  
APPLICANT: Gangolli, Esha A  
APPLICANT: Malyankar, Uriel M  
APPLICANT: Gunther, Erik  
APPLICANT: Smithson, Glennda  
APPLICANT: Millet, Isabelle  
APPLICANT: Gerlach, Valerie  
TITLE OF INVENTION: Proteins, Polynucleotides Encoding Them and Methods of  
FILE REFERENCE: 21402-221  
CURRENT APPLICATION NUMBER: US/10/023,634  
CURRENT FILING DATE: 2002-06-28  
PRIOR APPLICATION NUMBER: 60/256,025  
PRIOR FILING DATE: 2000-12-15  
PRIOR APPLICATION NUMBER: 60/265,163  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: 60/272,929  
PRIOR FILING DATE: 2001-03-02  
PRIOR APPLICATION NUMBER: 60/274,864  
PRIOR FILING DATE: 2001-03-09  
PRIOR APPLICATION NUMBER: 60/276,688  
PRIOR FILING DATE: 2001-03-16  
PRIOR APPLICATION NUMBER: 60/277,880  
PRIOR FILING DATE: 2001-03-22  
PRIOR APPLICATION NUMBER: 60/286,409  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/309,246  
PRIOR FILING DATE: 2001-07-31  
PRIOR APPLICATION NUMBER: 60/315,600  
PRIOR FILING DATE: 2001-08-29  
NUMBER OF SEQ ID NOS: 132  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 93  
LENGTH: 5636  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (3003)  
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the  
OTHER INFORMATION: specification.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (3041)  
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the  
OTHER INFORMATION: specification.  
FEATURE:  
NAME/KEY: VARIANT  
LOCATION: (3367)  
OTHER INFORMATION: Wherein Xaa is any amino acid as defined in the  
OTHER INFORMATION: specification.  
US-10-023-634-93

Query Match 71.6%; Score 2627.8; DB 14; Length 5636;  
Best Local Similarity 9.6%; Pred. No. 9.8e-30;  
Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;

QY 1 M-S-----I-----SS---D-----EV----- 8  
DB 1 MISWEVHTVFLFALLYSLAQDASPQSEIRAEFFEGASTLAFVDFVTGSMYDDLQVI 60  
QY 9 -----NF-LV----- 12  
DB 61 EGASKILETSLKPKPLNFALPHDPEIGPVTTITDPKFKQYELRELYVGGGDCPE 120  
QY 13 -----Y-----R---Y-----LQ----- 17

DB 121 MSIGAIIKIALEISLPGSFYIVFTDARSKDYRLTHEVLQIIQQKQSQVVFVLTGDCDDRTH 180  
QY 18 -----E-----SG-F-----S-H-S-----A-----F 26  
DB 181 IGYKVEEIIASTSSGOVFLDKKQVNEVLKQVVEEAQASKVHLLSTDHLEQAVNTWRIPF 240  
QY 27 -----T-----FG-----I-----E----- 31  
DB 241 DPSLKEVTVSLGSPSPMIEIRNPLGKLIKKGFGHELLNIHNSAKVNVNKEPEAGMWTVK 300  
QY 32 ---S-H-----I-----S-Q----- 36  
DB 301 TSSGRHSVRITGLSTIDFRAGFSRKPTLDFKTKVSRPVGIFTYVLLNTSGISTPARID 360  
QY 37 -----S-----NI-----N-----G----- 41  
DB 361 LLELLISGSSSLKTPVKYYPHRKPGYIWNISDFVPNEAFFLKVGTGYDKDDYLFQEVSS 420  
QY 42 -----A-----L-----VP-----P-----A 47  
DB 421 VSFSSIVDPAPKVTMPEKTPGYVLPQGPICPSVDLSLLPFTLSFVRNGVTGLGVDOYLKESA 480  
QY 48 -----A-----L-----I-----S-----IIQ-----KG----- 56  
DB 481 SVSLDIKAVTSLDEGEFYECIAVSAGTGRAQTFFDVSEPPVPIQVPPNNTVTTEGRAVLT 540  
QY 57 -----L-Q-----Y-----V-E----- 61  
DB 541 CLIIISAVDYNLTWQRNDVRVLAEPARITLANLSLELSKVKENDAGEVHCMSSEGSS 600  
QY 62 -A-----EVS-----NE----- 68  
DB 601 AASVFLTVQEPKVTVMKQNSFTGSGSEVSIMCSATGYPKPKIAVTNDMFIVGSHRYRM 660  
QY 69 ---DGLTF-----D-----G-----R-----P-----I-- 78  
DB 661 TSDGTLFIKNAAPKADAGIYGCILASNSAGTDKQNSTLRYTEAPKLMVQSELLVALGDITV 720  
QY 79 -E---S-----L-S-LI-----DA-----V----- 87  
DB 721 MECKTSGIPPPQVQKWFKGLDLELRPSTFLIIDPLGLLKIQETQDLQDAGDYTCVAINAGR 780  
QY 88 -----M-----P-DV-----VQ-----T----- 94  
DB 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRLLDNMFPSRPPS 840  
QY 95 -----R----- 95  
DB 841 VSSISQLRTGALFILNLWASDKGTICEAENQFGKIQSETTIVTGLVAPLIGISPSVAN 900  
QY 96 ---QO-----A-----Y---R-D-KL---AQ----- 105  
DB 901 VIEGQQLTLPCTLLAGNPPIERRWIKNSAMLLQNPYITVRSQGLSHIERVOLQDGGEYTC 960  
QY 106 -----QH-----A----- 108  
DB 961 VASNVAGTNKTTSVVVHVLPTIQHQQLISTIEGIPVTLPCCKASGNPKPSVIWKKGEL 1020  
QY 109 -----A-A-A-----A-A-A----- 114  
DB 1021 ISTSSAKFSAGADGSLYVVSPEGESGEYVCTATNTAGYAKRKVQLTVYVPRVFGDLRG 1080  
QY 115 -----A-----A----- 116  
DB 1081 LSODKPEIVSLAGEBVTLPCEVKSLLPPIITWAKETQLISPPSRHTFLPSCSMKITET 1140  
QY 117 -----ATN-----Q-----Q-GSA-K-----N----- 126  
DB 1141 RTSDSGMYLCVATNIAGNVTQAVKLVNHPVKIQRGPKHLKVQVQGVQVDIPCNAAQGTPLP 1200  
QY 127 -----GE---N-----T-A-N-G----- 133  
DB 1201 VITWSKGGSTMLVDGSHHVSNDGTLSIDQATPSDAGIYTCVATNIAGTDEBITLHVQE 1260

134 QY E E N GA H 139  
1261 Db PPTVEDLEPPYNTTFQERVANQRIEFPCKGTPKPTIKWLHNGRELTGREPGISIERG 1320  
140 QY T IA N HT DM 148  
1321 Db TLLVIASVTPYDNGEYICVAVNEAGTERKYNLKVHVPVVIKDBQVSNVLLNQLTNL 1380  
149 QY M E 150  
1381 Db FCEVEGTPSPINWYKDNVQVTESSITQVNNKGILKLFRAPEADAGRYSCRAINIAGTS 1440  
151 QY VD G DV E I P 158  
1441 Db QKYFNIDVLVPPTIIGTFNFKVSVLNRDVALEQCVKGTPFPDIHWFKDGKFLGDPN 1500  
159 QY SN K A V V 165  
1501 Db VELLDRGQVLHLKVARNDKGRYOCTVSNAGKQAKDIKLTIVIPPSIKGNVTTDISVL 1560  
166 QY RG H 168  
1561 Db INSLIKLECKTRGLPMPAITWYKDGQPIMSSQALVIDKQYLHI PRAQVSDSATYTCV 1620  
169 QY E S E VPI C A W 177  
1621 Db ANVAGTAEKSFHVDVYVPPMIEGNLATPLNKQVVIASHLTLECNAAGNPSPILTLKQGV 1680  
178 QY N P V 182  
1681 Db PVKANDFRIEAGGKLEIMSAQEI DRGQIVCATSVAGEKEIYKVDVLVPAIEGDE 1740  
183 QY LL VSGS G D S 191  
1741 Db TSVFVWNNLLELDCHVTGTPPTIMWLKQGLIDERDGFILLNGRKLVIQAQVSNT 1800  
192 QY TA R I W 196  
1801 Db GLYRCMAANTAGDHKKEFEVTVHVPPTIKSGLSERVVVVKYKVALQCIANGIPNPSITW 1860  
197 QY NL S E NS T S 204  
1861 Db LKDDQPVNTAGNLKIOSGSRVLQIAKTLEADAGRYTCVATNAAGETQOHIQLHVHPPS 1920  
205 QY G PTOL V L R 212  
1921 Db LEDAGKMLNETVLVSNPVQLECKAAGNPVPTWYKDNCLLSGSTMFLNRGQIIDIES 1980  
213 QY H C I 215  
1981 Db AQISDAGIYKCVAINSAGATELFYSLQVHVAPISGSGNNMVVVVNNPVRLCEARGIPA 2040  
216 QY R E G Q D 221  
2041 Db PSLTWLKDGPVSSFSNGLQVLSGRILALTSTQISDTGRYTCVAVNAAGEKQIDLRV 2100  
222 QY VPSN KD VT S L 231  
2101 Db YVPPNIMGEQNSVLISQAVELLQSDAIPPTTLTWLKDGHPLKPKGLSISENSRVLK 2160  
232 QY D W N S EG TLL 240  
2161 Db IEDAQVQDTGRYTCBATNVAGTEKKNVNNIWPVPPNIGGSDDELQTLTVIEGNLISLCE 2220  
241 QY AT GS Y D G F A 249  
2221 Db SSGIPPPNLIWKKGSPVLITDSMGRVRLSGRQLQISIAEKSDAALYSCVASNAGTAK 2280  
250 QY R I WTKG 256  
2281 Db KEYNLQVIRPTITNSGSHPTIIVTRCKSISLECEVOGIPPTVTTWKKOGLHIKAGV 2340

257 QY NL AS T L QH 265  
2341 Db EILDEGHILQKNHIVSDTGRVVCVAVVAGWTKKYDLSVHAPPSIIGNHRSPENISVV 2400  
266 QY K G PI FALK W 274  
2401 Db EKNSVLTCEASGIPLPSTTF KDGMPVLSLSNVRILSGRMLRLMOTTMEDAGQYTC 2458  
275 QY K K GN 279  
2459 Db VVRNAAAGEERKIFGLSVLPVPHIVGENTLEDVKVEKOSVTLTCEVTGNPVEITWHKOG 2518  
280 QY F I L S AGV 286  
2519 Db QPLQDEAHIIISGGRFLQITNVQVPHTCRYTCLASSPAGHKRSFSLNVFVSPTIAGVG 2578  
287 QY D K T T I W 293  
2579 Db SDGNPEDVTILNSPTSLVCEAYSYPATITFKDGTPLSNRINIRILFGRTLQILNAQ 2638  
294 QY D A H TG E 299  
2639 Db EDNAGRYSCVATNEAGEMIKHYEVKYIPPIINKGDLWGPGLSPKEVKIKVNTLTLECE 2698  
300 QY A K Q 303  
2699 Db AYAI PASLSWYKDGQPLKSDDDHVNIANGHTLQIKEAQISDTGRYTCVASNIAGEDEL 2758  
304 QY F P H 307  
2759 Db FVNIQVPPSFOKLWBEIGNMLDTGRNGEAKVNIINNPISLYCETNAAPPPTLTWYKDGHP 2818  
308 QY SA A LDV D 315  
2819 Db LTSSDKVLILPGGRVLQIPRAKVEDAGRYTCVAVNEAGEDSLQYDVRVLVPPILIEGANS 2878  
316 QY WQ SN 319  
2879 Db LPEEVTVLNKSALIECLSSGSPAPRNSWQKQPLLEDHKKFLSNRILQILNTQITD 2938  
320 QY NT FAS C S 326  
2939 Db IGRYVCVAENTAGSAXKYFNLNVHVPVPSVIGPKSENLTWVNNFISLTCEVSGFPDLS 2998  
327 QY T D MCI 331  
2999 Db WLKXQPIKLTNTLIVPGRTLQIIIRAKVSDGGEYTCIAINXAGESKKFSLTVVYVPS 3058  
332 QY H V C K L GO 338  
3059 Db IKDHSLSLVNVREGTSVLECESNAVPPVITWYKNGRMITESTHVEILADGOMLHI 3118  
339 QY D R P IK 343  
3119 Db KKAESDVTGOYVCRAINVAGRDDKNFHLNVVPPSIEGPEREVIVETISNPVTLTCDATG 3178  
344 QY T F Q H T 349  
3179 Db IPPPTIAWLKNYKRIENSLSLEVRILSGSKLQIARSQHSDSGNYTCIASNMEGKAQKY 3238  
350 QY N E V NA IKW D P T 360  
3239 Db FLISIQVPPSVAGAEIPSDSVLLGENVELVCCANGIPTPLIOWLKDGKPIASGETERIV 3298  
361 QY G NLL A S C S D D MTL 373  
3299 Db SANGSTLNIYGALTSDTGKTCVATNPAGEEDRIFNLNVVVTPIRGNKDEAKLMTYVD 3358  
374 QY K I W SM 378  
3359 Db TSINIECRXTGTPPPQINWLKNGPLPLSLSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418  
379 QY K Q DN CV 384



Db 481 SVSLDIKVTLSDEGVEICLAUSSACTGRAQTFDFVSEPPVIVQVNNVTVTPGERAVLT 540  
QY 57 -----L-Q-----Y-V-E-----61  
Db 541 CLIIISAVDYNLTWQRDRDVRLAEPARIRTLANLSLELKSVEKFNDAEGEYCHMVSSSEGGSS 600  
QY 62 -A-----EVS-----NE-----68  
Db 601 AASVFLTVQEPKVTWMPKNSQFTGSGSEVIMSATSATGPKPKIAVTNDMFIVGSHRYRM 660  
QY 69 --DGTILF-----G-----R-----P-----I--78  
Db 661 TSDGTILFKNAAPKDAIGYCLASNAGTDKQNSTLYRIEAPKLVVQSELLVALGDITV 720  
QY 79 -E-S-----L-S-LI-----DA-----V-----87  
Db 721 MECKTSGIPPPQVKFKGDLERFSTFLIIDPLGLLKIQETQDLDDAGDYTCVAINAAGR 780  
QY 88 -----M-----P-DV-----VQ-----T-----94  
Db 781 ATGKITLDVSGPPVFIQEPADVSMEIGSNVTLPCYVQGYPEPTIKWRRLDNMFIKRPFS 840  
QY 95 -----R-----OH-----A-----108  
Db 841 VSSISQIRTCALFILNWDKCTYICEAENQFGKIQSETTVTVTGLVAPLIGISPSVAN 900  
QY 96 ---QO-----A-----Y-R-D-KL-----AQ-----105  
Db 901 VIEGQQLTLPCTLLAGNPIPERWRKNSAMLLQNPYITVRSDSLHIERVQLODGEYTC 960  
QY 106 -----OH-----A-----108  
Db 961 VASVAGTNNKTSVVVHVLPITQHQOILSTTEGIPVTLPCASGNPKPSVWKKGEL 1020  
QY 109 -----A-A-A-----A-A-A-----114  
Db 1021 ISTSSAKFSAGAGSLVVSPEGESYVCTATNTAGYAKRVQVLTVYVRPVFGDLRG 1080  
QY 115 -----A-----A-----116  
Db 1081 LSQDKPVEISVLAGEEVLTPCEVKSLLPPIITWAKETQLISPPSPRHTFLPSSMKITET 1140  
QY 117 -----ATN-----Q-----Q-GSA-K-----N-----126  
Db 1141 RTSDSGYLCVATNIAGNVQAVKLVNHPVKIQRGPKHLKVQGVQGVDPICNAOQTPLP 1200  
QY 127 -----GE-----N-----T-A-N-G-----133  
Db 1201 VITWKGSTMLVDGEHHVSNPDGTLISIDQATPSDAGIYTCVATNIAGTDEBITLHVQE 1260  
QY 134 -----E-----E-N-----GA-----H-----139  
Db 1261 PPTVEDLEPPYNTTFOERVANORIEPPCPAKGTPTKTIKWLHNGRELGTGREPGISILEG 1320  
QY 140 T---IA-----N-----N-HT-----DM-----148  
Db 1321 TLLVIASVTPYDNGEYICVANEAGTTERKYNLKVHVPVVKKEQVSNVLLNQLTNL 1380  
QY 149 -----M-----E-----150  
Db 1381 FCEVEGTPSPIMWYKDNQVNTESSTIQTVNNKILKLFRATPEDAGRYSCAINIAGTS 1440  
QY 151 -----VD-----G-----DV-E-----I-----P-158  
Db 1441 QKYENIDVLPPTTIIGNFPKSVNLNRDVALEQVKGTFPPFDIHFWDKGLFLGDPN 1500  
QY 159 -----SN-----K-A-----V-----VL165  
Db 1501 VELLDRQVHLKVARNDKGRYQCTVSNAGKQAKDKLTIYIPPSIKGNVTTDISVL 1560  
QY 166 -----RG-----H-----168

Db 1561 INSILKLECKTRGLPMPAITWYKQOPIMSSSQALYIDKQYLHPIRAQVSDSATYVTCV 1620  
QY 169 -----E-S-----E-----VFI-----C-A-----W-----177  
Db 1621 ANVAGTAKSHFVDVYVPMIEGNLATPLNKOVVIAHSLTLECNAGNPSPLTLWLKQGV 1680  
QY 178 -----N-----P-V-SD-182  
Db 1681 PVKANDNFRIEAGGKKLEIMSAQEIDRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740  
QY 183 -----LL-----VSGS-----G-----D-----S--191  
Db 1741 TSFYFVWNNLLELDCHVTGSPPTIMMLKQQLIDRDFGFKILLNGRKLKLVIAQAQVSN 1800  
QY 192 -----TA-----R-----I-W196  
Db 1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSSGLSERVVVVKYKPVALQCIANGIPNPSITW 1860  
QY 197 -----NL-----S-----E-----NS-T-----S204  
Db 1861 LKDDQPVNTAQNLIKQSSGRVLQIAKTLELDAGRYTCVATNAAGETQOQIQLHVHEPPS 1920  
QY 205 -G-----PTOL-----V-----L-R-----212  
Db 1921 LEDAGKMLNETVLVSNPVOLECKAAGNPVPVITWYKDNCLSGSTMTFLNRQIIDIES 1980  
QY 213 -----H-----C-----I--215  
Db 1981 AQISDAGIYKCAINSAGATELFYSLQVHVAPSISGSNNVAVVNNPVRLCEARGIPA 2040  
QY 216 -----R-----E-G-----G-Q-D-----221  
Db 2041 PSLTWLKDGPVSSFNGLQVLSGGRILALTSTQISDTGRYTCVAVNAAGEKQORDILRV 2100  
QY 222 -VPSN-----KD-----VT---S-L-231  
Db 2101 YVPPNIMGEQNVSVLISOAVELLQSDAIPPTTLTWLKDGHPLLKPGLSISENSVLK 2160  
QY 232 -----D-----W-N-S-----EG---TLL--240  
Db 2161 IEDAQVQDGRYTCENTVAGTEKKNVNNVWPENIGSDELTLTQTVIEGNLISLCE 2220  
QY 241 AT-----GS-Y-D-----G-----F-----A-249  
Db 2221 SSGIPPPNLIWKKGSPVLTDMSGRVRIILSGRQLOISAIEKSDAALYSCVASNVAGTAK 2280  
QY 250 -----R-----I-----I-----WTKDG-----256  
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSIISLECEVQGIPTPTVTWMDKQGPLIKAKGV 2340  
QY 257 -----NL-AS-----T-----L-----GQH-----265  
Db 2341 EILDEGHILQLKNIHVSDDTGRYVCVAVNVAGTDDKYDLSVHAPPSIIGNHRSPEISVV 2400  
QY 266 -K-----G-PI-----PALK-W-----274  
Db 2401 EKNSVLTCEASGILPSTTF--KDGWPVSLNSVRILSGGRMLRMQTTMEDAQOYTC 2458  
QY 275 -----N-----K-K-----GN-----279  
Db 2459 VVRNAGEERKIFGLSVLPPHIVGENTLEDVVKKEKQSVTLTCEVTGPNVPBITWHKDG 2518  
QY 280 -----F--I-----L-----S---AGV-286  
Db 2519 QPLOEDEAHIIISGGRFLQITNVQVPHTRGYTCCLASSPAGHKRSFSLNVFVSPTIAGVG 2578  
QY 287 -D--K--T-----T-I-----I-W-----293  
Db 2579 SDGNPEDVTVILNSPTSLVCEAYSPPTATITWFKDGTPLSNRNRILPGRGTLOILNAQ 2638  
QY 294 -D-----A-----H-----TG-----E299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVKVIPIPIINKGDLWGFLSPKVEKIKVNNTLTLECE 2698



QY 300 A-----K-----Q-----Q----- 303  
Db 2699 AYAIASLSWYKDGOPKSDHVNIAANGHTLQIKEAQSITGRYTCVASNIAGEDEL 2758  
QY 304 P-----P-F-----H----- 307  
Db 2759 FDNVQVPPFQKWEIGNMLDTRNGEAKDVIINNPISLYCETNAAPPPTLTWYKDGHP 2818  
QY 308 --SA-----P-----A-----LDV-----D 315  
Db 2819 LTSSDKVILPGGRVLQIPRAKVEDAGRYTCVAVNEAGDSLOYDVRVLVPPPIEGANS 2878  
QY 316 -----WQ-----SN----- 319  
Db 2879 LPEEVLVNSALIECLSSGSPAPRNSWQKQDQPLEDDHFKFLNSGRILQILANTQID 2938  
QY 320 -----NT-----FAS--C-----S 326  
Db 2939 IGRYVCVAENTAGSAKYFNLVHVPVPSVIGPKSENLTVVVNNFISLTCEVSGFPPDLS 2998  
QY 327 -----T-----D-----MCI----- 331  
Db 2999 WLKNQOIKLNTLIVPGGRTLQIRAKVSDGEYTCIAINXAGESKKFSLTVVPPS 3058  
QY 332 --H-----V--C-----K-----L--GO----- 338  
Db 3059 IKDHSESLSVNVREGTSVLECESNAVPPVITWYKNGRMITBSTHVEILLADQMLHI 3118  
QY 339 -----D-----R-----P-IK----- 343  
Db 3119 KKAESVDTQYVCRAINVAGRDDKNFNLNVVPPSIEGPEREVIVETISNPVTLTCDATG 3178  
QY 344 -----T-----F-----Q--C-----H-----T----- 349  
Db 3179 IPPPTIAMLKNYKRIENSLSLEVRILSGSKLIQAKSHSDSGNYTCIASNMEGKAQKY 3238  
QY 350 -----N-E-V-NA-----IKW--D--P-----T----- 360  
Db 3239 FLISQVPPSVAGAEIPSDVSLLGENVELCVNANGIPTPLIQWLKDGKPIASGETERIV 3298  
QY 361 --G--NULL-A--S-----C--S-----D-----D-----MTL-- 373  
Db 3299 SANGSTLNIYALTSDTKYTCVATNPAGEEDRIFNLNVVTPITIRGNKDEAKLMTYVD 3358  
QY 374 -----K-----I-W-----SM----- 378  
Db 3359 TSINIECRXTGTPPPQINLKNGLPLPLSSHIRLLAAGQVIRIVRAQVSDVAVYTCVASN 3418  
QY 379 -----K--Q-----DN-----CV----- 384  
Db 3419 RAGVDNKHYNLQVAPPNNMDSMGTEETVLKGSSTSMACITDGTAPPSMAWLRDQOPLG 3478  
QY 385 --H-----D--L-----QA-----H-N-----KE 393  
Db 3479 LDAHLTVSTHGMVLQLLKAETEDSGKYTCIASNEAGEVSKFILKYLEPPHINGSEHEE 3538  
QY 394 I-----YT-----I--K--W----- 399  
Db 3539 ISVINNPULETCIASGIPAPKMTWMDKGRPLPQTDQVOTLGGGEVLRISTAQVEDTGRY 3598  
QY 400 -----SPTG-----P--GT-----N----- 407  
Db 3599 TCLASSPAGDDDDKEYLVRVHPVNPNTAGTDEPRDITVLRNQTLECKSDAVPPVITWLR 3658  
QY 408 N-----P-----N-----ANLM-LAS--A--S--F----- 420  
Db 3659 NGERLOATPRVILSGRVLQINNADLGTANTVCVASNIAGKTTREFILTVNVPNIK 3718  
QY 421 -----DSTV-----RL-W--D--V-----DR-----G--I----- 433  
Db 3719 GPQSLVILLNKSTVLECTIAEGVPTPRITWRKQDGLVAGNHNARYSILENGFLHQAHVTD 3778

QY 434 -----C-----I-----H-----TL-----T----- 439  
Db 3779 TGRYLCMATNAAGTDRRRIDLVHVPSPSIAPGPTNMTVIVNQTTLLACEATGIPKPSINW 3838  
QY 440 -K--H--Q-E-----P-V-----Y----- 446  
Db 3839 RKGHLLNVQDQNSYRLLSSGLVIFSFVDDTATYECTVTNGAGDDKRTVDLTVOVPP 3898  
QY 447 SVA-----F-S-----P-D----- 453  
Db 3899 SIADPTPLVTKHAPAVITCTASGVPPFSIHWTNGIRLLPRGQYRILSSGAIBLAT 3958  
QY 454 -----GRY-----L--ASG--S-- 461  
Db 3959 QLNHAGRYTCVARNAAGSAHRHVLHVEPPVIQOPSELHVLNNPILLPCEATCTPSP 4018  
QY 462 F-----D-----KCV----- 466  
Db 4019 FITWQKEGINVNTSGRNHVLPSGGLQISRAVREDAGTYMCVAQNPAGTALGKIKLVQV 4078  
QY 467 -----H-----I-W----- 469  
Db 4079 PPVISPFLKEYVIAVDKPTILSCREADGLPPPDI THWKDGRAIVESIRQVRVSSGSLQIAF 4138  
QY 470 -----N-----T-----Q----- 472  
Db 4139 VQPDAGHYTCMAANVAGSSSTSTKLTVHVPPIRSTEGHYTVNENSQAILPCVADGPT 4198  
QY 473 -----T-----G-----A-----L-VH 478  
Db 4199 PAINWKNVLLANLLGKYTABPYGELILENVLEDGTYTCVANNAAGEDTHTVSLTVH 4258  
QY 479 -----S-----Y-----R 481  
Db 4259 VLPPTTELPGDVSINKGBQLRSLCKATGIPKLTWTFNNIIPAHFDSVNGHSELVIER 4318  
QY 482 -----GT-----G-----IF-----E--V--C----- 490  
Db 4319 VSKEDSGTYVCTAENSVGFKATGTVYVKEPPVFKGDPSPNIEPLGGNAILNCEVKGDP 4378  
QY 491 --WN-----A-----AGD----- 496  
Db 4379 TPTIQNRKGVDBIEISHRIQLNGSLAIYGTVNEDAGDYTCVATNEAGVWVERSMTLQ 4438  
QY 497 -----K-V--G-----A-----SAS----- 503  
Db 4439 SPPIITLSPVETVINAGKILNCOATGEPQPTITWSROGHSISWDDRVNVLNNSLYTA 4498  
QY 504 -----D-----GSV-----C-V----- 509  
Db 4499 DAQKEDTSEFECVARNLMSVLVRVPVIVQVHGFSGQMSAWRACSVTCGKIQKRSRLCN 4558  
QY 510 --L-----DL--R-----K 514  
Db 4559 QPLPANGGKPCQGSLEMRNCONK 4582

## RESULT 9

US-10-464-368-69  
; Sequence 69, Application US/10464368  
; Publication NO. US20040023356A1  
; GENERAL INFORMATION:  
; APPLICANT: Krumlauf, Robb  
; APPLICANT: Ellies, Debra  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION  
; FILE REFERENCE: 40716-IP-017  
; CURRENT APPLICATION NUMBER: US/10/464,368  
; PRIOR FILING DATE: 2003-06-16  
; PRIOR APPLICATION NUMBER: 60/388,970  
; PRIOR FILING DATE: 2002-06-14  
; NUMBER OF SEQ ID NOS: 140  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 69

```

; LENGTH: 4599
; TYPE: PRT
; ORGANISM: MOUSE
US-10-464-368-69

Query Match      71.2%; Score 2612.9; DB 15; Length 4599;
Best Local Similarity 9.5%; Pred. No. 6.5e-30;
Matches 436; Conservative 55; Mismatches 17; Indels 4081; Gaps 381;

QY 1 MS-----I-----S-----4
Db 1 MSQALLAILTSLGLENVLIIGANQDQHLCDPGEFLCHDHVTCVSQWLCDGDPDPCD 60
QY 5 -SD-----EV-----N-----F-----10
Db 61 QSDSESLDTPCEEVEIKPLNHIACHGSSACVHLKLCNGVVDPCDPDFDEGHGQELLPS 120
QY 11 -L-----V-----Y-----R-----Y-L-Q-----17
Db 121 QQLNCOFKAMVFNATRCYCEDGFEVAEDGRSCKDQDECSIYICISQTCNKTYGSVAC 180
QY 18 -ESG-----F-----S-----H-----23
Db 181 VE-GYIMQSDNRSCVKHEPTDKAPMLLISSLETIELFYINGSKMTLSSANRNEIHTLD 239
QY 24 -----S-----S-----AF-T-----27
Db 240 FIYSEMICWIESRESSNQLKCOITKAGRLTDORINSLSQSPQVQWAFDMLTRNIYP 299
QY 28 -----F-G-----IES-H-----33
Db 300 VDHVSDFRIFVCFNENGVCVTLIESELNPKAIAADPIAGKLPFTDYGVPKIERCDLDM 359
QY 34 -I-----S-----Q-----36
Db 360 NRTRIVSKAEQPSALALDLNRLVYVDLYDYGVDYQGNRHTIVQGRQVHLYGI 419
QY 37 -----S-NI-----NG-----A 42
Db 420 TVFEDYLYATSDNFNIIRNFRNGTDIHSIIKMSARGIRTYQKRTQPTVRSHACEVDA 479
QY 43 -----42
Db 480 YGMPGGCHICLLSSSKTTCRCRTGPNMGSDGRCKRPKNELFLPYGKRGPIVRGMD 539
QY 43 L-----VP-----P-----A-A-LI-----S-II--OK-G-----L- 57
Db 540 LNTKIADECMPIENLVNPRALDFHAEANYIYFADTTSLIGRQKIDGTERETILKDDLD 599
QY 58 -Q-----Y-----V-----E-A-----E-----VS-IN-- 67
Db 600 NVEGIADVITGNLYWTDNGHRKTINVARLEKASQSRKTLLEGGMSPRAIVDPVNGWM 659
QY 68 -----ED-----G-TL-F-----D-----74
Db 660 YWTDWKEDKIDDSVGRLEKAWGVNRQVFVTSKMLWPNGLTLDFHTSLTYWCDAYYDHI 719
QY 75 -----G-R-----P-----I-----E-SL-----81
Db 720 EKVELNGTHRKVVYSGKELNHPFLSHGHVYFWFTDYMNGSIFQLDLMTNEVTLRHERA 779
QY 82 -SL-I-----D-----A-V-----87
Db 780 PLFGLQIYDPRKQGGNMCRINNGCGGTLCIAIPAGRVACADNQLLDENGTTCTFNPEE 839
QY 88 -----MP-----D-V-----V 92
Db 840 IRFHICKPGEFRCKNKHCIQARWKCDGDDCLDGSDEDSVTCFNHSCPDQFQCNNRCI 899
QY 93 -----QT-----R-Q-Q-----AY--R-----DK- 102
Db 900 PKRWLCDGANDCGSNEDESNQCTARTCOADQFSCGNRGRCIPTAWLCDREDDCGQDTDEV 959
```

```

QY 103 -----LAQ-----Q-----106
Db 960 ASCEPPTCEPLTQFTICKSGRCISNKNWCHDTHDDCGDRSDREVCCVHSCLDQDQRCSSGRCI 1019
QY 107 -H-A-----AAAA-A-----114
Db 1020 PGHWACDGDNDGDFSDETHINCTKEARSAGCIGNBFQCRPDGNCIPDLWRCDGEKDC 1079
QY 115 -----A-A-AT-----N-----Q-----120
Db 1080 EDGSDEKCGNGTIRLCDHKTKFSCRTGRCINNAWCDGDVDCDQSDDEEDCDLFCGPP 1139
QY 121 -----Q-----GSAK-----NG-----127
Db 1140 KYPCANDTSVCLQPEKLCNGRKDCPDGSDGDLCDCECSLNNCGSNHSCVWVPGRGIVCSC 1199
QY 128 -E-----N-T-----A-----N-GE-----E 135
Db 1200 PEGHQLKDNRTCEIVDYCASHLRCSQVCEQQQHMVKCSCYEGWALGTGDSCTSVDSFE 1259
QY 136 -----TD-----M-----M-----E-----150
Db 1260 AFIIISIRHEIRIDLHKGDYSLLVPLGRNTTALDFHFNQSLLYWTDVVEDRIYRGKLE 1319
QY 137 -G-A-----H-----T-----IA-N-----N-----H 145
Db 1320 SGVSAIEVVVEHGLATPEGLTVDWIAIGNIYWDNSLDQIEVSKLDGSLRATLIAGAMEH 1379
QY 146 -----DG-D-VEI-----P-----S-----N-KA-----162
Db 1380 PRAIALDPYGLTFWTDMDANFPRIESASMSGAGRKTIYKDMKTGAWPNGLTVDHFERRI 1439
QY 151 V-----V-----L-----RG-----H-----168
Db 1440 VMTDARSDAIYSAFYDGTNNMIEIRGHEVLSHPFVAVSLYGVYVWTDWNTLAKANKWT 1499
QY 163 -----VV-----L-----RG-----H-----168
Db 1500 QGNVSVIOKTSQAQPDQIYHPSRQPAQPNPCAANEGRGPCSHLCLINHRSAACAPHL 1559
QY 169 -----E-----SE-----V-----172
Db 1560 MKLSSDKTCYEMKFKFLYARRSEIRGVDIDNPYNFITFTVDPDIDDVAVIDFADSEER 1619
QY 173 -----FI-----174
Db 1620 LYWTDIKTQITRAFINGTGLTVISRDIOQIRGLAVDWVRNLYWISSEFDETOINVAR 1679
QY 175 -----C-A-----W-----N-----PV--S 181
Db 1680 LDGSLKTSIIHGIDKPCQCLAAHPVRGKLYWTDGNTINMANMDGNSKILFQNKQEPVGLS 1739
QY 182 -D-----LL-V-SGSG-----D-----S-----TAR-I-----W--NL-- 198
Db 1740 IDYVENKLYWISSGNGTINRCNLDOGNLEVIEMKEELTKATALITMDKKLWADONLAQ 1799
QY 199 -----S-----S-----E--NS-----TS-----204
Db 1800 LGTCNKRDRGNPSILANKTSIGVVMKVYDKEAQGSNSCOVNNGCSQCLPTSETTRIC 1859
QY 205 -----G-----P-----206
Db 1860 MCTVGYVLQKNRMSQGISFLMYSVHEGIRGIRPLEPRDKVDALMPISGAFAVADIFHA 1919
QY 207 -----T-----Q-----LV--L-R-----H-----C- 214
Db 1920 ENDTIYWTMDGLNKISRAKRQDTWKEDVVVNGLRVREGIAVDWIAIGNIYWTDHGFNLIEV 1979
QY 215 -----I-----R-----E-G-----GO-----D-----V-- 222
Db 1980 ARLNGSFRVVISQGLDQPRSTAVHEKGFLEWTEWQVPCIGKARLDGSEKVMIVSVGI 2039
QY 223 --P--S--NK-----D-----V--T-----S-LD-----W- 233
```



QY 472 --Q--T-G--AL--V--H--S-----Y 480  
 DB 4256 YCQNGGTCTPSTLGRPTCICALGFTGPNCGKAVCEDSCHNGGSCVVTAGNPYCHCQADY 4315  
 QY 481 --R--G--T--G--GG-----485  
 DB 4316 TGRCOYYVCHYCVNSSECTGNDGVEVCPTRYEGPKCEIDKVCRHGHCHCIKNKN 4375  
 QY 486 --IF--E--VC--W-----491  
 DB 4376 EDIFCNCTNGKIASSCOLCDGYCYNGGTCTQOLDPETSIPVCVCSTNWSGTQCPERPAPKSSK 4435  
 QY 492 -----491  
 DB 4436 SEHISTRSAIIVPLVLLVTLVTLVGLVWCKRRTKTIRROPIINGINVEIGNPSY 4495  
 QY 492 NA--AG--D-K--VG--A--SA-----SD--G--505  
 DB 4496 NMYEVHDHSDGGLLEPSPFIDPVKRYICGGSSAFKLPHTAPPYVINSDLKGLPTFGPT 4555  
 QY 506 --S--V-----C-----VLDLRK 514  
 DB 4556 NYSNPVYAKLYMDGQCRNSLASV-DERK 4583

## RESULT 10

US-10-464-368-70  
 ; Sequence 70, Application US/10464368  
 ; Publication No. US20040023356A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Krumlauf, Robb  
 ; APPLICANT: Ellies, Debra  
 ; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR REGULATING BONE DEPOSITION  
 ; FILE REFERENCE: 40716-IP-017  
 ; CURRENT APPLICATION NUMBER: US/10/464,368  
 ; CURRENT FILING DATE: 2003-06-16  
 ; PRIOR APPLICATION NUMBER: 60/388,970  
 ; PRIOR FILING DATE: 2002-06-14  
 ; NUMBER OF SEQ ID NOS: 140  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 70  
 ; LENGTH: 4599  
 ; TYPE: PRT  
 ; ORGANISM: MOUSE  
 US-10-464-368-70

Query Match 71.2%; Score 2612.9; DB 15; Length 4599;  
 Best Local Similarity 9.5%; Pred. No. 6.5e-30;  
 Matches 436; Conservative 55; Mismatches 17; Indels 4081; Gaps 381;

QY 1 MS-----I-----S-----4  
 DB 1 MSQLLAILTLGLLPNAEVLIVGANQDHLCDPGEPLCHDHVTCVSQWLCDDGDDPCPD 60  
 QY 5 SD-----EV-----N-----F-----10  
 DB 61 QSDSLTCTPEVEIKPLNHIACHGSSACVHLSKLCNGVVDGPDGDEGHCOELLPS 120  
 QY 11 --L--V--Y--R--Y--L--Q-----17  
 DB 121 QQLNCFKCAVRNATRCYCEDGEVAEDGRSCKQDECSIYIGCSQTKNTYGSYACSC 180  
 QY 18 --SG-----F-----S-----H-----23  
 DB 181 VE-GYIMQSDNRCKVXHEPTDKAPMLLISSLETIELFYNGSKWTLSSANRNEIHTLD 239  
 QY 24 -----S-----AF--T-----27  
 DB 240 FIYSEMICWIESRESSNQLKCGQITKAGRLTDORIINSQSFQNVQMAFDLWLRNIYF 299  
 QY 28 -----F-G-----IES--H-----33

DB 300 VDHVSDRI FVCNFNGSVCVTLIESLHNPKAIAADPIAGKLFFTDYGNVPKIERCDLDM 359  
 QY 34 ---I---S---Q-----36  
 DB 360 NRTRIVYSKAEQPSALALDLNRLVYVVDLYLDYVGVVDYQGNKRHTIVQGRQVHLGYI 419  
 QY 37 ---S--NI--NG-----A 42  
 DB 420 TVPEDYLYATSSDNFIIRINRFNGTDIHSIIKMESARGIRTYQKRTQPTVRSACEVDA 479  
 QY 43 -----42  
 DB 480 YGMPGGGSHICLLSSSYKTRTCRCRTGFNMGSDGRSKPKNELFLFYKGRPGIVRGMD 539  
 QY 43 L-----VP-----P-----S-II--OK--G-----L- 57  
 DB 540 LNTKIADCEMIPENLVNPRALDFHAEANYIYFADTTSLIGRQKIDGTERFETILKDDLD 599  
 QY 58 --Q-----Y-----V--E-A-----E-----VS-IN-- 67  
 DB 600 NYEGIAVDWIGNNLYWTDNGHKTINVARLEKASQSRKTLLEGGMHSHPRAI VVDPVNGWM 659  
 QY 68 -----ED-----G-TL-P-----D-- 74  
 DB 660 YWTDWKEDKIDDSVGRIBKAWMDGVNRQVFTSKMLWPNGLTLDFTSTLYWCDAIYDHI 719  
 QY 75 -----G-R-----P-----I-----E-SL----- 81  
 DB 720 EKVFLNGTHRKVVYSGKELNHPFGLSHGNVYVFTDYNGSIFQLDLMTNEVTLRHERA 779  
 QY 82 --SL-I-----D-----A-V-----87  
 DB 780 PLFGLQIYDPRKQDGNMCRINNGCGTCLAI PAGRVACADNQLLDENGTTCTFNPEE 839  
 QY 88 -----MP-----D-V-----V 92  
 DB 840 IRFHICKPGEFRCKNKGICQARWKCDGDDCDLGDSDSVTCFNHSCPDQKQCNRCI 999  
 QY 93 -----QT--R--Q--Q-----AV--R-----DK- 102  
 DB 900 PKRWLCDGANDCGSNEDESNOICTARTCOADQFSCGNGRCIPTAWLCREDDCGDTDEV 959  
 QY 103 -----LAQ-----Q-----106  
 DB 960 ASCERPTCEPLTQFICKSGRCISNKHWCDDDDCGDRSDEVCVHSCLDQFRCSGRCI 1019  
 QY 107 --H-A-----AAAA--A-----114  
 DB 1020 PGHWACDGDNDGDFSEDETHINCTKEARSAGCIGNEFQCRPDGNCIPDLWRCDGEKDC 1079  
 QY 115 -----A-A--AT--N-----Q-----120  
 DB 1080 EDGSEKNGGTIRLCDHKTFCSTGRGINNAWVCDGDVDCEDSDDEDCDSFLCGPP 1139  
 QY 121 -----Q-----GSAK-----NG-----127  
 DB 1140 KYPCANDTSVCLQPEKLCNGRKCDPGSDEGLDCDECSLNNGCSNHCSVVPGRGIVCSC 1199  
 QY 128 --E--N-T-----A--NGE-----E 135  
 DB 1200 PEGHQLKDNRTCEIYDYCASHLRCSQVCEQKHVYKSCYEGWALGTGDESCTSVDSFE 1259  
 QY 136 -----N-----136  
 DB 1260 APTIFSRHRIERIDLHKGYSLLVPLGLRNTIALDFHFNQSLLYWTDVVEDRIVRGKISE 1319  
 QY 137 --G--A-----H-----T--IA-N-----N-----H 145  
 DB 1320 SGGSVAIEVVVEHGLATPEGLTVDWIAIGNIYWDNSLDQIEVSKDGLSRATLIAGAMEH 1379  
 QY 146 -----TD-----M-----M-----E-- 150  
 DB 1380 PRAIALDPYRGLFWTDWDANFPRIESSASMGAGRTIYKDMKTGAWPNGLTVDHFERRI 1439

151 V-----DG-D-VEI-----P-----S-----N-----KA-----162  
1440 VMTDARSALISAFYDGTNMIIEIRGHEVLSHPFAVLSYGSVEYVWTDWRTINTLAKANKWT 1499  
163 -----VV-----L-----RG-----H-----168  
1500 GONVSVIQTSAOPFDLQIYHPSQPAPNCPAANEGRGPCSHLCLINHNRSAAACPHL 1559  
169 -----E-----SE-----V-----172  
1560 MKLSSDKICYEMKKELLARRSEIRGVDINDPYNFITAFTVPDIDDVAVIDFDASEER 1619  
173 -----FI-----174  
1620 LYWTDIKTQITRAFINGTLETVISRDIQSIRGLAVDWVRNLYWISSEFDETOINVAR 1679  
175 -----C-A-----W-----N-----PV--S 181  
1680 LGSLKTSIIHGIDKPQCLAAHPVRGKLYWTDGNTINMANMDSGSKILLFQNOKEFVGLS 1739  
182 -D-----LL-V-SGSG-----D-----S-----TAR-I-----W-----NL-- 198  
1740 IDIVENKLYWISSGNGTIKRCNLDDGNLEVIEMKEELTKATALTIMDKKLWADONLAQ 1799  
199 -----S-----E-----NS-----TS-----204  
1800 LGTCNKRDGRNPISILNKTSIGVVHMKVYDKEAOGSNCQVNNGCSQLCLPTSETTRIC 1859  
205 -----G-----P-----206  
1860 MCTVGYLQKNRMSQIGIESFLMYSVHEGIRGIPLEPRDKVDALMPIGSAAPAVGIDFHA 1919  
207 -----T-----Q-----LV--L-R-----H-----C- 214  
1920 ENDTIYDMLGNKISRAKRDQTKEDVVTNGLGRVEGIAVDWIAGNIYWDTHGFNLIEV 1979  
215 -----I-----R-----E-G-----GQ-----D-----V-- 222  
1980 ARLNGSFRVYIISQGLDQPSIAVHPKGFPLFTWENGQVPCIGKARLDGSEKVMIVSVGI 2039  
223 --P--S-----NK-----D-----V-----T-----S-LD-----W- 233  
2040 TWPNGISIDYEENKLYWCDAARSKIERIDLDGTANREVLISGNSVDLFSVAVGAYIYMS 2099  
234 -----N-S-----E-GT-----238  
2100 DRAHANGSVRRGHKNDAETETVMTGTGLVNLKEIKIFNRVREKGTNVCAKENGCCOOLCL 2159  
239 -----LLA-----T-----G-----S-----244  
2160 YRGNSRRTACAHGYLAGDGVTCRLRHEGYLLYSGRILKLSIHLSDETNLNSPVRPYENPN 2219  
245 Y-----D-----G-----F--A-----R-I-----251  
2220 YFKNIILAFDYNORREGTNRIFYSDAHFNGIQLIKDNWEDROVIVENVGSVEGLAYHRA 2279  
252 -----W-----TK--D-----G-----NL-----258  
2280 WDTLYWTSSTSSITRHTVDTQTRPGAIDREAVITMSDEDDHPHYALDECONLMFWTNWNE 2339  
259 -----A-----ST-----LG-----Q-H- 265  
2340 QHPSIMRATLTGNAHVVVSTDLITPGLTIIDHRAEKLIFSDGSLGKIERCEYDGSQRHV 2399  
266 --K-GP-----IP-ALKW-----NK-----K-----277  
2400 IVKSGPGTFLSLAVDSYIFWS--DWGRRAILRSNKYTGGETKILRSIDIPHOPGIIAVAN 2458  
278 -----G-----N-----F-- 280  
2459 DTNSCBLSPCALLGGCHDLCLLTPDGRVNCSCRGDRVLLANNRCVTKNSSCNIYSEFEC 2518  
281 -----I-----L-----S--AG-----V-----286  
2519 GNGDCYDYVVLTCDDGI PHCKDKSDEKLLYCNRSRSGFKPCYNRRVCVPHGKLCODGTDCG 2578  
287 -----D-K-----T-----T-I-----291  
2579 DSSDELUDCKVSTCSTVEFRCADGTCIPRSARCQNQNMDCSDASDEKGNNTDCTHIFYKLG 2638  
292 -----I-----W-----D-----AH-----T 297  
2639 KSTGFRRCNSTRSLCVLPSPWICDGSNDGDYSDBLKCPVQNHKKEENYFCPSGRCLINT 2698  
298 -----G-E-----A-----K-----301  
2699 WVCDOGKDCEDGLDELHCDSSCSWNQFACSVKVCISKHMITCDGEDDCGDSLDESISCGA 2758  
302 -----Q-----Q-----F-----304  
2759 VTCADMFCQSHACVPOHWCDBRDCPDGSDSSAGCAPNNTCDENAFMCHNKVCI 2818  
305 P--F--H-----S-----AP-----ALD-----VD--WQ-----S-- 318  
2819 PKQFVCHDDDCDGSDEFLQCGYRQCGPEEPRCA--DGRCLVNTLWQCDGDFDCPDSSDE 2877  
319 --N-----NT--FASC-----STD-MC-I-----HV--C--KL 336  
2878 APINPRCSAEHSCNSFFM--CKNGRCIPS--DGLCDIRDGCGSDGTNCHINECLSKKI 2935  
337 -G-QD-R-PI-----KT-----F--Q-----G-----347  
2936 SGCSQDQDLPSYKCKWPGFQLKDGKTCVDIDECSSGFPSCSQCCINTYGYKCHCAE 2995  
348 -----H-----H-----T-----N-----E 351  
2996 GYETQDPNPGCRSLSDDEPFLILADQHEIRKISTDGSNTLLKQGLNNVIALDFDYREE 3055  
352 -----V-NA-----I-K--W-D-----358  
3056 FIYWDSSRPNGSRINRMCLNGSDIKVHVHTA VPNALAVDWIGKLYWSDEKRIIEVSK 3115  
359 -----PT-----GNLL-ASC-----S- 368  
3116 LINGLYTVLVSRLKFPRLDLSLDPNAGNLYWIDCCBYPHIGRVMGDTGNSVVIETKISR 3175  
369 -----D-----M-----TL-K-I-W-----376  
3176 PHALITIDYVNHRLYWADEHNHIFESNMDSHRHKVPNQDIPGVIALTLFEDYIYWDGKTK 3235  
377 --S--MK-----Q-D-----N-----C-----V 384  
3236 SLRSHVKTSGADRSLSLNSWHAITDIQVYHSVRQPDVSKHLCTVNNGGCSHLCLLSPGKT 3295  
385 H-----D--L-----387  
3296 HTACPTNPFYLAADNRNLTCLSNCTASQFRCKTDKCIPIFWKCDTVDVDCGSDSDPDCEP 3355  
388 -----Q-----A-----H-----NKE-I-----394  
3356 KCQPGRFQGTGLCALPAFICDGENCGNSDELNCDDTHVCLAGQFKCTKNKKCIPVNL 3415  
395 -----Y-----T-----I-K-W--SP-----T 402  
3416 CNGQDDCGDEDEKDCPENSCLSPDYFQCKTTHKICISKLWVCDDEPDCADASDEANCCKKT 3475  
403 -GP-----G-----T-N-----N-P-----409  
3476 CGPHEFCQKNNCIPDHWRCNDQNSDSDNCKPOTCTLKDFLCSNGDCVSSRFWCD 3535  
410 -----N-----A-----N-----412  
3536 GBFDCADGSEKNCETSCSKDQFQCSNGQCLSAKWCDGHEDCKYGEDEKNCBPAPVCS 3595  
413 --LMLA-----SAS-----F-----420

Db 3596 SSEYMCASGCLSLASLKCNGEPCDVGSDMDVIECKEDQFQCKNKNKAYCIPRWLDCGI 3655  
QY 421 ---D-S---T---V-R---L-W---D-V---429  
Db 3656 YDCVDSDEETCGRGSGICRDDEFLCNLSLKLHFWVCDGEDDCGSDNSDEAPDMCVKFLC 3715  
QY 430 ---DR---GI---CI 435  
Db 3716 PPTPRYCRNDRICTLOLEKICNGINDCGDSDBEHCGKLSLXKPKCKDEFTCSNRNCI 3775  
QY 436 ---HT---437  
Db 3776 PMLQCDLSDCGSDGDEQCLKTPTEHTCENNGNPGDDAYCNQIKTSVFCRCKPGFQR 3835  
QY 438 ---L-T---K---440  
Db 3836 NMKGRECADLNECLLFGICSHHCLNTRGSYKVCQDNQFQEKNNCSIAKSGEDQALYIAND 3895  
QY 441 ---HQ---E---P-V-Y---446  
Db 3896 TDILGFVYPNYSGGHQISHVEHNSRITGMVHYQRNVIVWSTQFNPGGIFYRMIDARE 3955  
QY 447 ---S---VA---FS---451  
Db 3956 KQANGSLICPEFKRPDIADVWAGNVYVTDHSMHWFYSYTHWTSLSYSINVQQLNG 4015  
QY 452 ---P---DG---454  
Db 4016 PNCTRLLTNAGBPYATVNPVKRGMWYTVIGDHSHEEAAMDGLRLRVLQKQLRPTG 4075  
QY 455 ---R-Y---LA---S---GS---461  
Db 4076 LTVDFGERIYADFELSIITGSVLYDGSFVSSVSKQGLLHPHRIDVPEDYIYGAPKN 4135  
QY 462 ---F---DK---CV 466  
Db 4136 GIFRVQFGHGSVEVLALGVDTKTSILVSHRYQLNLPNPCLDLSCDFCLINPNSGATCI 4195  
QY 467 ---H---I---W---N-T-471  
Db 4196 CPEGKYMNGTCHDDSLDDSKLTCBNGRCILNEKGLRCHCWPSYSGRCVNHCSN 4255  
QY 472 ---O---T-G---AL---V---H---S---Y 480  
Db 4256 YCQNGGTCTPSTLGRPTCICALGFTGPNCGKAVCEDSCHNGSCVVTAGNPQYCHCOADY 4315  
QY 481 ---R---G---T---GG---485  
Db 4316 TGRQCQYVCHYCVNSECTIGNDGSVEVCPTRYEGPKCEIDKCVRCGGHCILNKDN 4375  
QY 486 ---IF---E---VC---W---491  
Db 4376 EDIFCNCTNGKIASQCQCDGVCYNGGTCLDPETSIPTVCVTNWSGTQCPERPAPKSSK 4435  
QY 492 ---NA---AG---D-K-VG---A---SA---SD---G---505  
Db 4496 NYVEVDHSDGGLLEPSFMDIPVKSRYIGGSSAFKLPHTAPPYILNSDLKGLPTFGPT 4555  
QY 506 ---S-V---C---VLDLRK 514  
Db 4556 NYSNPVYAKLYMDQNCNRNSLASV-DERK 4583

## RESULT 11

US-10-120-801-74  
; Sequence 74, Application US/10120801  
; Publication No. US20030203843A1  
; GENERAL INFORMATION:

; APPLICANT: Pena, Carol  
; APPLICANT: Guo, Xiaojia  
; APPLICANT: Shimkets, Richard  
; APPLICANT: Padigaru, Muralidhara  
; APPLICANT: Kekuda, Ramesh  
; APPLICANT: Spytek, Kimberly  
; APPLICANT: Mehraban, Fuad  
; APPLICANT: Topper, James N.  
; APPLICANT: Malyankar, Uriel  
; APPLICANT: Wasserman, Scott  
; APPLICANT: Edinger, Shlomit  
; APPLICANT: Smithson, Glennda  
; APPLICANT: Gunther, Erik  
; APPLICANT: Komuves, Laszlo  
; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same  
; FILE REFERENCE: 21402-340  
; CURRENT APPLICATION NUMBER: US/10/120,801  
; CURRENT FILING DATE: 2002-04-11  
; PRIOR APPLICATION NUMBER: 60/285748  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: 60/286068  
; PRIOR FILING DATE: 2001-04-24  
; PRIOR APPLICATION NUMBER: 60/286292  
; PRIOR FILING DATE: 2001-04-25  
; PRIOR APPLICATION NUMBER: 60/288334  
; PRIOR FILING DATE: 2001-05-03  
; PRIOR APPLICATION NUMBER: 60/291241  
; PRIOR FILING DATE: 2001-05-16  
; PRIOR APPLICATION NUMBER: 60/322284  
; PRIOR FILING DATE: 2001-09-14  
; PRIOR APPLICATION NUMBER: 60/285609  
; PRIOR FILING DATE: 2001-04-20  
; NUMBER OF SEQ ID NOS: 155  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 74  
; LENGTH: 5175  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-120-801-74

Query Match 71.2%; Score 2612.1; DB 14; Length 5175;  
Best Local Similarity 9.2%; Pred. No. 1.2e-29;  
Matches 446; Conservative 53; Mismatches 12; Indels 4339; Gaps 386;  
QY 1 M-----SI-----SS-----D-----E-----7  
Db 1 MGRSPSLYGLVGLLLALLTTCSSVNDKNDPTGKSLAFVFDITGSMFDDLQVREGAAK 60  
QY 8 ---V---N---F-L-VY-----13  
Db 61 IFKTVMAQREKLIYNIYMPFHPDYLGEIINTDSTYFMQLSKVYVHGGDCPEKLTG 120  
QY 14 ---R-Y---L-QE--SG--F-----SHSAF-T 27  
Db 121 ILKALQISLPSSFIYVFTDARSKDYHLEDEVLATIQEKSSVVFVMTGDCGNRTHPGFT 180  
QY 28 ---FG-I---E---S-----H-ISQ-----S-NI-----39  
Db 181 YEKIAAASFGQVPHLEKSDVSTVLEYVRHAVKQKVHLMAYEARERGTVSRNIPVDKHL 240  
QY 40 ---N---N-----GA-----LV---P-P-----46  
Db 241 ELTISLSDKDDSDNLDIVLRDPGRTVDKRLYSKEGGTIDLKNVLRKLDSPSGVTV 300  
QY 47 ---A-----A-----LI-S-I-----52  
Db 301 NTSRLKHTIRVFGHGAVDPKYGFASRPDLRIELARPPVNLQDTVLLINMTGLIPGTV 360  
QY 53 --I-----Q-----KGL-----O-Y-----VE-----AE-63  
Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGFPVFKGLFFVRVQGVDEDEYFEMRIAPT 420  
QY 64 ---V-----S-I-----N-----ED-----GTILF-----73

Db	421	AIGSVIVGPRAFMSPIHOEFVGRDLNLSCTVESASAYTIYWKITGEDIIIGPLFYHNTD	480	Db	1499	DKTSWKTSDDKRKLHVFKAKITDTSYGYKCVARNAAGEGSKSFQVEVIVPLNLDSEKYYKKK	1558
Qy	74	-----D-----G-----R-----P-I-----E-----	79	Qy	217	-----EG-----G-----Q-----D-----V-P-----SN-----	225
Db	481	TSWITPELSLKDAGEYECRVISNNGYSVKTRVETRESPELFGVRNVSVPLGEAFLH	540	Db	1559	VPAKEGEEVTLGCPVSGFPVPQINWVVDGTVVVEPGKKYKATLSNDGLTLHPDSVSVKQE	1618
Qy	80	-----S-----L-----SL-I-----DA-----	86	Qy	226	-----N-SE-----G-----KQVT-S-L-----D-----W	233
Db	541	CSTRSAGEVEIRWTRYGATVFNCPNTPNPTNGTLKIHVTRADAGVYECMARNAGMST	600	Db	1619	GNHCVASQKGNILIDVELSVLAVPIVGEDDNLEVLFGKDISLSCDLQTESDDKTTFW	1678
Qy	87	-----VM-----P-DV-----V-----	92	Qy	234	-----N-SE-----G-----TL-----L-----	240
Db	601	RKMRDIMEPPSVKVTPODYFNMREBGNLSCEAMGDPKPEVHWYFKGRHLLNDYKYQG	660	Db	1679	SINGESDRPDNVQIPSGHRLYITDAPENNGKYMCRVTNAGKAERTLTLDVLEPPVF	1738
Qy	93	-----T-----R-----Q-OAYRD-----KLA-----Q-----	105	Qy	241	-----A-----TG-----SY-D-----	246
Db	661	QDSKFLYIRDATHDEGTTECRAMSAGQA-RDITDLMLATPKVEIIONKMWVGRGDRV	719	Db	1739	VEPVFEANQKLGNNPIILQCVTGNPKPTVIWKIDGNDVDSWLFDELSLLRIEKLGT	1798
Qy	106	-----H-----Q-H-----A-----AAA-----AA-----	113	Qy	247	-----G-----F-----A-----R-----I-----	251
Db	720	SPECKTIRCKPHKIRWFKNGDLIKPDDYIKINEQLHMGAKDEADAGAYSCVENMAG	779	Db	1799	KSAQISCTAENKAGTASRDFFIQNTAAPTFKNEGQETIFRESEITITLDCPVSLGDFQIT	1858
Qy	114	-----AAAA-----T-----N-----Q-----Q-GS-----A-KN-----	126	Qy	252	-----W-----T-----T-----KD-----	255
Db	780	KDVQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNP	839	Db	1859	WMKQGLPLTENDAIFTLNTRLTILNANRDHEDIYTCVANNTAGQVSKDFVWVQVLPKI	1918
Qy	127	-----GEN-----T-----A-N-----	132	Qy	256	-----GN-----L-----A-----S-----	260
Db	840	RYTLADG-NLLITDAQIEDQOFTCIARTYQOQSOTLMTVGLVSPVLGHVPPEBOL	898	Db	1919	KNAVTLNEGEEIILTCDAEGNPTPTAKWDFNQDLPKEAVFVNNHTVVNNVTKYH	1978
Qy	133	-----G-----E-----E-----N-----G-----	137	Qy	261	-----L-----GQ-----H-----K-----	266
Db	899	IEQDLTLCVVVLTGPKPSIWIWIKDDKPVVEGPTIKIEGGSLRLRGNPKDEKYYTC	958	Db	1979	TGVYKCYATNKVGQAVKTINVHRTKPRFESGLTESELTVNLTRSTLECDVDDAIGVGI	2038
Qy	138	-----A-----H-----TI-----A-N-NH-----	145	Qy	267	-----G-P-----I-----F-----	270
Db	959	IAPSPAGNGLHINVQLIKKPEVPYKPEGIVFKPTISGMDEKHVAVNSHTDVLDEGF	1018	Db	2039	SWTVNGKPLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVTNEAGVATKTNLFVQVP	2098
Qy	146	-----G-----E-----E-----T-----	146	Qy	271	-----AL-----K-----W-----	274
Db	1019	AIPCVSVSGTPPIITWYLDGRPIITNSRDTFTVADNTLIVRKADKSYSVGYTCQATNSAG	1078	Db	2099	PTIVNEGGEYTVIENNSVLPCVETGKPNPVVTWKDGRPVGLKSVQVLSGQQPKIVH	2158
Qy	147	-----M-M-----E-----VD-----GD-----V-----	155	Qy	275	-----N-----K-----KG-----	278
Db	1079	DNEQKTIIRMTPIISPQSSFNWVDDLFTIPCDVYGDPKPVIITWLLDDKPFTEGVN	1138	Db	2159	ABIAHKSYYICAKNDVGTAEISFDVDIITRPMIOKGIKNIVTAIKGALPFKCPIDDDK	2218
Qy	156	-----I-----PS-----N-K-A-----VYL	165	Qy	279	NP-----IL-----S-----AG-----	285
Db	1139	EDGSLTIPWNEAHRGFTTCHQAQAGNDRTRVTLTVHTTPTINAEQKIALQNDIVL	1198	Db	2219	NFKGOIILWLRNYQPIDLEADARIITLSNDRRLTILNVTENDEGGQYSCRVKNDAGENSFD	2278
Qy	166	-----R-----	167	Qy	286	-----V-----D-KT-----TI-----I-W-D-A-H-----	296
Db	1199	ECPAKALPPVRLWYEGEKIDSLIPHITREDGALVLQNVKLENTGVFVQVSNLAGED	1258	Db	2279	FRATVLVPTIIMLDKDKKTAVEHSTVTLSCPATGKPEPDITWFKDGAHIENTADII	2338
Qy	168	-----HE-----SEV-----F-I-C-A-----WNP-V-----	180	Qy	297	-----T-----G-----EA-----K-----Q	302
Db	1259	SLSYTLTVHEKPIIIEVPGVDVVKGTIEIPCRATGVEVIRTNWKNKIDLMKDEKPF	1318	Db	2339	PNGELNGKLIKTRIEKGDAKYTCADNSAGSVEQDVNVVITIPKIEKOGIPSDYESQ	2398
Qy	181	S-D-L-L-V-S-----G-----SG-----D-----S-----TA-----	193	Qy	303	Q-----	303
Db	1319	SVDNLTGRIYEADKNDIGNYCNVVTNEAGTSQMTTHVDVQEBPIILPSTQTNNTAVGD	1378	Db	2399	QNERVVISCFVYARPPAKITWLKAGKPLQSDKFVKTSAHQKLYLPKLRETOSSKYTCIA	2458
Qy	194	RI-----W-----NL-----S-ENST-----S-----	204	Qy	304	-----F-----P-F-----H-----S-----	308
Db	1379	RVELKCYEASPPASVTFWFRGIAIGTDTKGYVVEDSGTLVLIQOSASVEDATIYTKASNP	1438	Db	2459	TNEAGTKRDFKVMVLVAPSFDPENIVRITVNSGNPSTLHCPAKSPFTITWLKDGNA	2518
Qy	205	-G-----P-TO-----L-V-----L-----	211	Qy	309	-----A-----	310
Db	1439	AGKAENLQVTVIASDDIKDPDVVTOESIKESHPSLYCPVFNPLPOISWYLNKPLID	1498	Db	2519	IEPNDRYVFDAGRLQISKTEGSDQGRYTCIATNSVSGDDELNTLEVIIPPIDGERRE	2578
Qy	212	-----R-H-----CI-R-----	216	Qy	311	-----A-----L-D-----VD-WQ-----S-N-----	319
				Db	2579	AVAVIEGFSSELFCDSNSTGVDEWQKGLTINQDTLRGDSFIQIPSSGKKMSFLSARKS	2638

QY 320 ---N---T---FA---S---C---S-T---327  
Db 2639 DSGRYTCIVRNPAEARKLDFFAVNDPPSISDELSSANIQTIVYYPVEINCVSQSPHP 2698  
QY 328 ---DM---C---I-331  
Db 2699 KVVWLFDDKLEPDSAAAYELTNNGETLKIVRSQVEHAGTYTCEAONNVGKARKDFLVRVT 2758  
QY 332 ---H---V---C---KL---G---337  
Db 2759 APHFKEKEBEVVARVGDTMLLTCNAESSVPLSVYWHADHESVQNGVITSKYAANEKTL 2818  
QY 338 ---Q---C---DR---PI---K---343  
Db 2819 NVTNIQLDDGFGYCTAVNEAGITKFKFLIVITIETPYFLDQKLYPIILGKRLTLDCSAT 2878  
QY 344 ---T---F---A---345  
Db 2879 GTPPPTILFMKDGKRLNESDEVDIIGSTLVINDPQKEVEGRYTCIAENKAGRSEKDMWVE 2938  
QY 346 ---Q---G-H---348  
Db 2939 VLLPPKLSKEWINVEVQAGDPLTECPIDETSGVHITWSRQFGKGQOLDMRAQSSDKSK 2998  
QY 349 ---T---N---E---VN---A---354  
Db 2999 LYIMQATPEADSYCIANDAGABAVFOVTNTPPKIFGDSFSTTEIVADTTLBPCR 3058  
QY 355 ---IKW---P---GN---D 358  
Db 3059 TEGIPPEISWFLDGRKPILEMPGVYTKQGLSLRLIDNIKENQGRYTCVAENKAGRAEQD 3118  
QY 359 ---P---T---GN---M---362  
Db 3119 TYVEISEPPRVVMASEVMRVEGQRTIRCEVEGNEPVPVWNLKDGEPYTSLLQFSTKL 3178  
QY 363 ---L---LA---SC---S---D---D---M---371  
Db 3179 SYLHLRETTLADGGTYTCIATNKAGESQTTTDEVILVPPRIEDEERVLOGKEGNTYVHC 3238  
QY 372 ---L---T---L---373  
Db 3239 QVTGRPVYVTKRNGKEIQPNVLHNRATRADEKYSKCIASNEAGTAVADFLIDVFT 3298  
QY 374 ---KI---W---S---M-K 379  
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGGRPFNMDNIILSPRGDTLMILK 3358  
QY 380 ---QD---N---C---383  
Db 3359 AQRFDGGLYTCVATNSYGDSEQDFKNVYTKPYIDETIDOTPKAVAGGELILKCPVLGNP 3418  
QY 384 ---V---H---DI---Q---A---H-NKE-I 394  
Db 3419 TPTVTWKRGDDAVPNDNRHTIVNNYDLKINSVTTEDAGQVSCIAVNEAGNLTHYAAEVI 3478  
QY 395 ---Y---TI-K---W---SP---401  
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLFSISWFRGDKFVYLYDRYSISFDGSHIT 3538  
QY 402 ---T---T---402  
Db 3539 INKAKLSDGKYTCRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYLECPI 3598  
QY 403 -G-P---G---T---NN---408  
Db 3599 SGIPQPDVITKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRVTCATNRRGKASHD 3658  
QY 409 ---P---409  
Db 3659 FSLDVLSPPEFDIHGTQPTIKREGDTITLTCPIKLAEDTADQVMDVSWTKOSRALDGDLT 3718

QY 410 -N---A---N---L---ML---A-SAS---F---D---421  
Db 3719 DNVDISDDGRKLTISOASLENAGLYTCIALNRAGEASLEFKVEILSPPIDISRNDVQPO 3778  
QY 422 ---S---T---VR---L---W---427  
Db 3779 VAVNQPTIMRCVTHGPPFSIKWLKNGKEVTDENIRIVEQOGVQLILRTSDSHAGKWSC 3838  
QY 428 ---DV---D---R---G---I---C---434  
Db 3839 VAENDAGVKELEMLVDVFTPPVSVKSDNPRIKALGETITLFCNASGNPYPQLKWKAGGSL 3898  
QY 435 ---I-H---T---437  
Db 3899 IFDSPGARISLKGARLDIPLHKKTDVGYTQCALNAAGTSEASVVDVLVPPEINRDGI 3958  
QY 438 ---LT---K---H---Q---442  
Db 3959 DMSPLPAQOGLTLOCLAQKGPVQMRWTLNGTALTHTPGITVASDSTFIQINNVSLS 4018  
QY 443 ---E---PV---Y---S---447  
Db 4019 KGVTCTAENVAGSDNLMYNDVWQAPVISNGGTQKQVIEGELAVIECLVEGYPAPQVSWL 4078  
QY 448 ---V---A---449  
Db 4079 RGNRVETGVQVRYVTDGRMLTIIEARSLDSGIYLCSATNEAGSAQOAYTLEVLVSPKI 4138  
QY 450 ---FS---P---DG---454  
Db 4139 ITSTPGVLTSPSSKFSFLCAVRGYPDPIISWTLNGNDIKDGENGTIGADGTLHIEKAE 4198  
QY 455 -R-Y---L---A---SG---460  
Db 4199 ERHLIYECTAKNDAGADTLEFFVQTIIVAPKISTSGNRYINGSEGTETVIKCEIESESEF 4258  
QY 461 ---S---F-D-K---C---465  
Db 4259 SMSKNGVPLLPNNLIFSEDYKLIKILSTRSDQGEYSCTAANKAGNATQKTNLVGVAP 4318  
QY 466 ---VH---I-W---NT---471  
Db 4319 KIMERPTQVHKGDQVTLWCESAGVQPAITWYKDNELLTNTGVDETATTKKSVIFSS 4378  
QY 472 ---O---TG---474  
Db 4379 ISPSQAGVYTCAEKNVWASTEEDIDLIVMIPPEVPEVMNVSTNPRQTVFLSCNATGIPE 4438  
QY 475 ---A---L---V---H---478  
Db 4439 PVISMRDSNIAIONNEKYQILGTTLAIERNVLPDDDFYHCIAKSDAGQKIATKRLIVNK 4498  
QY 479 ---S-YR-G-482  
Db 4499 PSDRPAPIWVECKEKGPKKTEYMDIRGDPDDNPOLLPWKQVEDSSLSNGSIAYRCMPGP 4558  
QY 483 ---T---G---I---FE---488  
Db 4559 RSRVTLLHAAPQFIVKPKNTAAIGAIVELRCSAAGPPHPHTITWAKDGKLIEDSKFEIA 4618  
QY 489 ---V-C-490  
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKSSNQKNVAVITCYE 4678  
QY 491 ---W---N-A---493  
Db 4679 RNOAYSRLGTWBYNGVPMKPNLAGIHFMMNGSLIVLDTSSLKEGDLLEYTCVKVRNRHS 4738  
QY 494 ---A-G---DKV---G---AS---A---502  
Db 4739 IPHLTSAFEGPEVKTIDKRVNNGSVVLDCVETSDPLTTHVVVTKNQKMLDDDAIYV 4798  
QY 503 ---S---D-GSV-CV---LD---L---R---K 514





Db 1679 SINGSEDRPNVQIPSDGRLIYITAKPENNGKYMCRVTSNAGKAERTILTLDVLEPPVF 1738  
QY 241 -A- -TG- -SY- -D- - 246  
Db 1739 VBPVFEANOKLIGNPFILOQCVTGNPKPTVIWKIDGNDVDKSWLFDLSLIRIEKLTG 1798  
QY 247 -G- -F- -A- -R- - 251  
Db 1799 KSAQISCTAENKAGTASRDFIIONIAAPTAKNEGDOETIFRSESETITLDCPVSLGDFQIT 1858  
QY 252 W- -T- -KD- - 255  
Db 1859 WKQGLPLTENDAIFLIDNTRLTILNANRDHEDIYTCVANNTAGQVSKDFVVVQVLPKI 1918  
QY 256 -GN- -L- -A- -S- - 260  
Db 1919 KNAVTLLEINEGEIEILTCAEGNPTPTAKWDFNOGDLPEKAEVFNANNHTVVVNNVTKYH 1978  
QY 261 T- -L- -Q- -H- -K- - 266  
Db 1979 TGVKCYATNKVQAVKTNVHVRTKPRFESGLTESELTVNLTRSLTECDVDDAIGVGI 2038  
QY 267 -G- -P- -I- -F- - 270  
Db 2039 SWTVNGKPFPLAETDGVQTLAGRFLHIVSAKTDHGSYACTVTNEAGVATKTNLFQVVP 2098  
QY 271 -AL- -K- -W- - 274  
Db 2099 PTIVNEGGEYTVIENNSLVPCEVTGKPNPVVWTGDRPVGDLKSVQLSEGOQPKIVH 2158  
QY 275 -N- -K- - 278  
Db 2159 AEIAHKSYSYCMKANDVGTAEISFDVDIIITRPMIOGKINIVTAIKGALPFPKCPIDDDK 2218  
QY 279 NF- -IL- -S- -AG- - 285  
Db 2219 NFKQIILWNLNYQIDLEADARITRLSNDRLTILNVNTEDEGOYSCRVKNDAGENSFD 2278  
QY 286 -V- -D- -KT- -TI- -I- -W- -D- -A- -H- - 296  
Db 2279 FKATVLVPTIIMLDKDKNTAVEHSTVTLSCPATGKPEPDITWFKDGEAIIHENIADII 2338  
QY 297 -T- -G- -EA- -K- -Q- - 302  
Db 2339 PNGELNGNQLKIRIKEGDAGKYTCEDNSAGSVEQDVNVNVTIPKIEKDGIPSDYESQ 2398  
QY 303 Q- - 303  
Db 2399 QNERVVISCPYARPPAKITWLKAGPLQSDKFVKTSANGQKLYLPKLRETDSSKYTCIA 2458  
QY 304 -F- -P- -H- -S- - 308  
Db 2459 TNEAGTKDRDFKVSMLVAPSFDEPNIVRRITVNSGNPSTLHCPAKGSPSTITWLKDGNA 2518  
QY 309 -A- -P- - 310  
Db 2519 IEPNDRYVFFDAGQLQISKTEGSDQGRYTCIATNSVGSDDLENTLEVIIPPIDGERRE 2578  
QY 311 -A- -L- -D- -VD- -WQ- -S- -N- - 319  
Db 2579 AVAVIEGFSSELCFDSNSTGVVDEWQDGLTINQDTLRGDSFIQIPSSGKKMSFLSARKS 2638  
QY 320 -N- -T- -FA- -S- -C- -S- -T- - 327  
Db 2639 DSGRYTCIVRNPAGEARKLFDFAVNDPPSISDELSSANIQTIVPYVPEINCVVSGSPHP 2698  
QY 328 -DM- -C- - 331  
Db 2699 KYVWLFDDKPLEPDSAAVELTNNGETLKIIVRSQVEHAGTYTCEAQNNVGKARKDFLVRVT 2758  
QY 332 -H- -V- -C- -KL- -G- - 337  
Db 2759 APPHFEKEREBWARVGDMLLTCAESSVFLSSVYVWHAHDESQVNGVITSKYAANEKTL 2818

QY 338 -Q- -DR- -PI- -K- - 343  
Db 2819 NVTNIQLDDEGGYCTAVNEAGITKKFKLIVETPYFLDQOKLPIILGKRLTLDCSAT 2878  
QY 344 -T- -F- - 345  
Db 2879 GTPPTILLPMKDGKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGRSEKDMMVE 2938  
QY 346 -Q- -G- -H- - 348  
Db 2939 VLLPPKLSKEWINVEQAGDPLTLECPIDTSGVHITWSRQFGKQGLDMRAQSSDSK 2998  
QY 349 -T- -N- -E- -VN- -A- - 354  
Db 2999 LYIMQATPEDADSYSCIAVNDAGGAFAVQVTVNTPPKIFGDSFSFTEIVAADTLEIPCR 3058  
QY 355 -IKW- - 358  
Db 3059 TEGIPPEISWFLDGKPILEMPGVTVYKQGLSLRIDNIKPNQEGRYTCVAENKAGRAEQD 3118  
QY 359 -P- -T- -GN- - 362  
Db 3119 TYVEISEPPRVVMASEWVRVVEGRQTTIRCEVFGNPEPVVNMKDGEPYTSLLQSTKL 3178  
QY 363 -L- -LA- -SC- -S- -D- -D- -M- - 371  
Db 3179 SYLHLRETTLADGGTYTCIATNKAGESQTTTDEVLVPPRIEDEERVLOGEGNTYVHC 3238  
QY 372 -T- -L- - 373  
Db 3239 QVTGRPVVYTWKRGKEIQFNPVLHNRATRADEKGYSCIASNEAGTAVADFLLDVFT 3298  
QY 374 -KI- -W- -S- -M- -K- - 379  
Db 3299 KPTFETHETFNIVAGESAKIECKIDGHPKPTISWLKGRPFNMMDNIILSPRGDTLMILK 3358  
QY 380 -OD- -N- -C- - 383  
Db 3359 AQRFDGLYTCVATNSYGDSEQDFKNVYTKPYIDETIDOTPKAVAGGEIILKCPVLGNP 3418  
QY 384 -V- -H- -DL- -Q- -A- -H- -NKE- -I- - 394  
Db 3419 TPTVTWKRGDDAVPNDSRHTIVNNYDLKINSVTTEDAGQVSCIAVNEAGNLATTHYAAEVI 3478  
QY 395 -Y- -TI- -K- -W- -SP- - 401  
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVYLYDRYSISPDGSHIT 3538  
QY 402 -T- - 402  
Db 3539 INKAKLSGGKYTCRASNEAGTSDIDLILKILVPPKIDKSNIIGNPLAIVARTIYLECPI 3598  
QY 403 -G- -P- -G- -T- -NN- - 408  
Db 3599 SGIPQPDVITWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCIATNRGGKASHD 3658  
QY 409 -P- - 409  
Db 3659 FSLDVLSPPEFDIHGTOPTIKREGDTITLTCPIKLAEDIADQVMDVSWTKDSRALDGLT 3718  
QY 410 -N- -A- -N- -L- -ML- -A- -SAS- -F- -D- - 421  
Db 3719 DNVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKEILSPPVVIDISRNDVQPO 3778  
QY 422 -S- -T- -VR- -L- -W- - 427  
Db 3779 VAVNQPTIMRCAVTHGHPFSPISIKWLKNGKEVTDENIRIVEBQGVQLTLRTDSHAGKMSC 3838  
QY 428 -DV- -D- -R- -G- -I- -C- - 434  
Db 3839 VAENDAGVKEMLVDFVTPPVSVVSKSDNFIKALGETITLFCNASGNPYQLKWAKGSL 3898

QY 435 -----I-H-----T-----437  
Db 3899 IFDPSGARISLKGARLDIFHLKKTVDGYDTQALNAAGTSEASVSDVLVPPEINRDGI 3958  
QY 438 -----L-T-----K-----H-----Q-----442  
Db 3959 DMSPLPQAQSLTLQCLAQKQVPQMRWTLNGTALTHTSTPGITVADSTFIQINNVSLSLSD 4018  
QY 443 -----E-----P-----V-----Y-----S-----447  
Db 4019 KGYYTCAENAVAGSDNLMYNDVVQAPVINSNGGTVKQVIEGELAVIECLVEGYAPQVSWL 4078  
QY 448 -----V-----A-----449  
Db 4079 RGNRVETGVGVYVTDGRMLTIIEARSLDGIYLCSATNEAGSAQAQVTLVLSPKI 4138  
QY 450 -----FS-----P-----DG-----454  
Db 4139 ITSTPGVLTPSGSKFSLPCAVRGYPDPPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4198  
QY 455 -R-Y-----L-----A-----SG-----460  
Db 4199 ERHLIYECTAKNDAGADTLEFPVQTVAPKISTISGRNYINGSEGTETVKEBIESESSEF 4258  
QY 461 -----S-----F-D-----K-----C-----465  
Db 4259 SMSKNGVPLLPNNLIFSEDYKLIKILSTRLSDOGEYSCTAANKAGNATKTNLVGVAP 4318  
QY 466 -----VH-----I-W-----NT-----471  
Db 4319 KIMERPRQVHVHKGQDQVTLWCEASGVPOPAITWYKDNELLTNTGVDETATTKKSVIFS 4378  
QY 472 -----Q-----TG-----474  
Db 4379 ISPSQAGVYTKAENWVASTEEDIDLIMIPPEVPERMVNPNRQTVFLSCNATGIPE 4438  
QY 475 -----A-----L-----V-----H-----478  
Db 4439 PVISWNRDSNIAIQNNKEYQILGTTLAIRNVLDDGDFYHCIAKSDAGQKIATKRLIVNK 4498  
QY 479 -----S-----YR-----G-----482  
Db 4499 PSRPAPIWECDEKPKKTEYIMIDRGTPDDNPQLPEWDESSLNGSIAYRCWPGP 4558  
QY 483 -----T-----G-----G-----I-----FE-----488  
Db 4559 RSRVTLLHAAPOFIVKPKNTTAAIGAVELRCSAAGPPHTTITWAKOGLIEDSKFEIA 4618  
QY 489 -----V-----C-----490  
Db 4619 YSHLKVTLNSTSDGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKSSNQKNVAVITCYE 4678  
QY 491 -----W-----N-A-----493  
Db 4679 RNQAYGRGLTWETNGVPMFKNLAGIHFMMNGSLVILDTSSLEKGDLELYTCKVRNRRHS 4738  
QY 494 -----A-G-----DKV-----G-----AS-----A-----502  
Db 4739 IPHLTSAFEGVPEVKTIKQVENVNGSDVLDCEVTSDELTHVVWTKNDQKMLDDDAIYV 4798  
QY 503 -----S-----D-GSV-----CV-----LD-----L-----R-----K 514  
Db 4799 LPNNSLVLLNVEKYDEG-VYKVASNSIGKAFDDTQNLNVGSSRREAYK 4847

## RESULT 13

US-10-369-493-6861

; Sequence 6861, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; PRIOR FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 6861

; LENGTH: 5175

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

; US-10-369-493-6861

## Query Match

71.2%; Score 2612.1; DB 14; Length 5175;

Best Local Similarity 9.2%; Pred. No. 1.2e-29;

Matches 446; Conservative 53; Mismatches 12; Indels 4339; Gaps 386;

QY 1 M-----SI-----SS-----D-----E-----7  
Db 1 MGRSPSNLYGVGLLLALLATTCSSVNDKNDPTGKSLAFVFDITGSMFDDLQVREGAAK 60  
QY 8 ----V-----N-----F-----L-----V-----13  
Db 61 IFKTWAQREKLIYINMVPHPDPYLGEINTDSTYFMRQLSKVYVHGGDCPEKTLTG 120  
QY 14 -----R-Y-----L-----OE-----SG-----F-----SHSAF-T 27  
Db 121 ILKALQISLPSSFYVFTDARSKDHYLEDEVLTNIQEKSSVVFVMTGDCGNRTHPGFRT 180  
QY 28 -----FG-I-----E-----S-----H-ISO-----S-NI-----39  
Db 181 YEKIAAASFGQVPHLEKSDVSTVLEYRHAVKQKVLHMYEARERGTVSRNTPVDKHL 240  
QY 40 -----N-----GA-----LV-----P-P-----46  
Db 241 ELTISLSDGKDDNDLIVLRDPGRTVDKRLYSKEGGTIDLKNVKLIRLKDPSPGWT 300  
QY 47 -----A-----A-----LI-S-I-----52  
Db 301 NTNSRLKHTIRVFGHGAVDKYGFAASPLDRPLRIELARPPVNLQDTYLLINMTGLIP 360  
QY 53 -I-----Q-----KGL-----Q-Y-----VE-----AE- 63  
Db 361 GBIDLVDYHGHSLYKAVASPHRTNPMYFAGPVFPKGLFFVRVQGYDEDNFMRIAPT 420  
QY 64 -----V-----S-I-----N-----ED-----GTLF-----73  
Db 421 AIGSVIVGGPRAPMSPIHQEFVGRDNLNLSCTVESASAYTIYWKGTEDIIGGLFVH 480  
QY 74 -----D-----G-----R-----P-I-----E-----79  
Db 481 TSVWTIPELSKLDAGEYECRVISNNGNSVYKTRVETRESPEIFGVNRVNSVPLGEAFL 540  
QY 80 -----S-----L-----SL-I-----DA-----86  
Db 541 CSTRSAGEVEIRWTRYGATVFNPNTERNTPTNGTLKIHVTRADAGVYECMARNAGMST 600  
QY 87 -----VM-----P-DV-----Q-----V-----92  
Db 601 RQRLDIMEPPSVKVTPODVYFNMREGVNLSCAMGDPKPEVHWYFKGRLHLLNDYKYQVG 660  
QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----Q-----105  
Db 661 QDSKFLYIRDTATHDEGTVECRAMSGAQA-RDITDMLATPPKVEIIQNKMMVGRGDRV 719  
QY 106 -----Q-H-----A-----AAA-----AA- 113  
Db 720 SPEKTIKPKPKIRWFKNGKDLKPKDDYIKINEGQLHMGAKORDAGAYSCVGENMAG 779  
QY 114 -AAAA-----T-----N-----Q-----Q-GS-----A-KN- 126

Db 780 KOQVANLSVGRPTTIESPHTVRNIEQVTLQCLAVGIPPPEIEWQKGNVLLATLNP 839  
QY 127 ---GEN---T--A-N----- 132  
Db 840 RYQLADG-NLLITDAQIEDQGFCTIARNTYQQSQSQTLLMVTGLVSPVLGHVPPPEQL 898  
QY 133 ---G---EE---N-G--- 137  
Db 899 IEGQDUTLSCVVLTGPKPSIVWIKDDKPVEEGFTIKIEGGSLRLRGNGPKDEKYTC 958  
QY 138 ---A---H---TI---A--N-NH----- 145  
Db 959 IAVSPAGNLTLLHINVQLIKKPEFVYKPEGGIVFKPTISGMDEKXHAVVNSTHVDLGEGF 1018  
QY 146 -----T----- 146  
Db 1019 AIPCVVSGTPPPIITWLDGRPTPNSRDFVTADNTLIVRKADKSGVGYTCQATNSAG 1078  
QY 147 D---M---E---VD---GD-----V- 155  
Db 1079 DNEQKTTIRIMTPTMISPGQSFMMVVDLFTIPCDVYGDPKFVITWLLDDKPFTEGVN 1138  
QY 156 E---I---PS---N--K-A---VVL 165  
Db 1139 EDGSLTIPVNEAHRGFTTCHAQNAAGNDRTRVTLVHTTPTINAENQEKIALQNDIVL 1198  
QY 166 -----R-----G-- 167  
Db 1199 ECPAKALPPVRLWTYEGEKIDSLIPHITREDGALVQNVKLENTGVFCVQVNSLAGED 1258  
QY 168 ---HE---SEV---F--I-C-A---WNP-V- 180  
Db 1259 SLSYTLTVHEKPKIIEVPGVVDWVGFTIEIPCRATGVPEVIRTNKNGIDLKMDEKXF 1318  
QY 181 S-D-L-L-V-S---G---SG---D---S---TA--- 193  
Db 1319 SVDNLGTLRIYEAADKDNIGNYCVVINEAGTSQWTHVDVQEPPIILPSTQNTNTAVGD 1378  
QY 194 RI-----W-----NL-----S-ENST---S- 204  
Db 1379 RVELKCYVEASPASVTWFRGIAIGTDTKGYVVESDGLVIOQSASVEDATIYTCRASNP 1438  
QY 205 -G-----P---TQ---L---V---L-- 211  
Db 1439 AGKAEANLQVTVIASPDIKDPDVVTOESIKESHPSFLYCPVFNPLPQISWYLNKPLID 1498  
QY 212 ---R-H---CI-R--- 216  
Db 1499 DKTSWKTDDKRLHVFPAKITDSGVYKCVARNAAGEGSKSFQVEVIVPLNLDSEKYYKK 1558  
QY 217 ---EG---G---Q---D---V-P-----SN----- 225  
Db 1559 VFAKEGEVTLGCPVSGFPVQINWVVDGTWVEPGKYKGATLSNDGLTLHFDVSVSVKQE 1618  
QY 226 ---KQVT-S-L-L---D---W 233  
Db 1619 GNYHCVAQSKNILDIDVELSVLAVPIVGBDDNLEVLFGKDILSLSCDLQTESDDKTTFW 1678  
QY 234 -N-SE-----G-----TL-L- 240  
Db 1679 SINGESDRPDNVQIPSDGHLRYITDAKPENNGKYMCRVNTNSAGKAERTLTLDVLEPPVF 1738  
QY 241 ---A---TG-----SY-D----- 246  
Db 1739 VEPVEANQKLIIGNPIILQCVTGNPKPTVIWKIDGNDVDKSWLFDSELSLLRIEKLJTG 1798  
QY 247 ---G---F---A-----R-----I- 251  
Db 1799 KSAQISCTAENKAGTASRDFQIONIAAPTFKNEGQDQETIFRESEITITLDCPVSLGDFQIT 1858  
QY 252 W-----T-----KD----- 255

Db 1859 WMKQGLPLTENDAIFTLDNTRLTILNANRDHEDIYTCVANNTAGVSKDFDVVVQVLPKI 1918  
QY 256 ---GN---L--A-----S--- 260  
Db 1919 KNAVTLTLEINEGEEIILTCDAEGNPTPTAKWDFNOGDLPEAFVNNHTVVNNVTKYH 1978  
QY 261 T---L---GQ---H--K----- 266  
Db 1979 TGYYKYATNKVGQAVKTINVHRKPRFESGLTSELTVNLTRSTLECDVDDAIGVGI 2038  
QY 267 ---G-P-----I-----F--- 270  
Db 2039 SMTVNGKPLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVTNEAGVATKTFNLFVQVP 2098  
QY 271 ---AL---K---W----- 274  
Db 2099 PTIVNEGGTYTVIENNSLVLPCEVTGKPNPVVTWKDGRPVGDLKSVQVLSGQQPKIVH 2158  
QY 275 ---N---K-----KG----- 278  
Db 2159 ABLAHKGSVIMAKNDVGTABEISFDVDIITRPMIOKGIKNIVTAIKGGALPFKCPIDDDK 2218  
QY 279 NF-----IL-----S-----AG----- 285  
Db 2219 NFKGQIILWRNYQPIDLEAEDARITRLSNDRLTILNVTENDEGOQYSCRVKNDAGENSFD 2278  
QY 286 ---V---D-KT---TI---I-W--D-A-H----- 296  
Db 2279 FKATVLVPFTIIMLDKDNKNTAVEHSTVTLSCPATGKPEPDITWFKDGEAHIENIADII 2338  
QY 297 ---T---G---EA-----K-----Q 302  
Db 2339 PNGELNGQLKTRIKEGDAGKYTCADNSAGSVEQDVNVNVTIIPKIEKDGPSPDESQ 2398  
QY 303 Q----- 303  
Db 2399 QNERVVISCPVYARPPAKITWLKAGKQLSDRFVKTSAWGOKLYLFLKJRETDSSKYTCIA 2458  
QY 304 ---F---P-F---H-----S----- 308  
Db 2459 TNEAGTKRDKFKVSMVLVAPSFDEPNIVRRITVNSGNPSTLHCPAKSPSTIITWLDGNA 2518  
QY 309 ---A----- 310  
Db 2519 IEBNDRYVFDAGRQIQISKTEGSDQGYTCIATNSVSGDDLENTLEVIIPPIDGERRE 2578  
QY 311 --A---L--D---VD--WQ---S-N----- 319  
Db 2579 AVAVIEGFSSELPFCDNSGTGVDVEMQKDLITINQDTLRGDSFIQIPSSGKKMSFLSARKS 2638  
QY 320 ---N---T---FA-----S-----C--S-T--- 327  
Db 2639 DSGRYTCIVRNPAAGEARKLDFAVNDPPPSISDELSSANTQITVIVPYVPVINCVSUSGPH 2698  
QY 328 ---DM-----C-----I- 331  
Db 2699 KVTWLFDDKPLEPDSAAVELTNNGETLKIVRSQVEHAGTVTCEAONNVGKARKDFLVRVT 2758  
QY 332 ---H---V---C---KL-----G----- 337  
Db 2759 APPHFEKEREVVARVGDITMLITCNAESSVPLSSVYWHADSVQNGVITSKYAANEKTL 2818  
QY 338 ---Q-----DR---PI--K----- 343  
Db 2819 NVTNIQLDDGEGYCTAVNEAGITKKPFKLVITETPYFDQOKLYFLILGKRLTLDCSAT 2878  
QY 344 ---T---F----- 345  
Db 2879 GTPPPILFMKCKRLNESDEVDIIGSTLVIDNPQKEVEGYCTIAENKAGRSEKMDMWE 2938  
QY 346 ---Q-----G-H----- 348  
Db 2939 VLLPPKLSKEWINVEVQAGDPLTLECIPTSGVHITWRSQFGKDGOLDMRAOSSSKSK 2998

QY 349 ---T---N---E---VN---A--- 354  
Db 2999 LYIWOATPEDADSYCIANDAGAEAVQVNTTPPKIRGDSFSTTEIVADTLEIPCR 3058  
QY 355 ---IKW--- 358  
Db 3059 TEGIPPEISWFLDKPILFEMPGVYKQGLSLRIDNIKNQOGRYTCVAENKAGRAEQD 3118  
QY 359 ---P---T---GN--- 362  
Db 3119 TYVEISEPRVMASEVMRVEGRQTTIRCEVFGNPEPVVNLKDGEPYTSLLQFSTKL 3178  
QY 363 ---L---LA---SC---D---D---M--- 371  
Db 3179 SYLHLRETTLDGGTTCIATNAGESQTTTDEVLVPPRIEDEERVLOKEGNTYVHC 3238  
QY 372 ---T---L--- 373  
Db 3239 QVTGRVPVYVWKRNGKEIQFNPVLHNRATRADEKYSYCIASNEAGTAVADFLIDVFT 3298  
QY 374 ---KI---W---S---M---K 379  
Db 3299 KPTFETHETTFNIVESAKIECKIDGHPKPTISWLKGRPFNMNDIILSPRGDTMLK 3358  
QY 380 ---QD---N---C--- 383  
Db 3359 AQRFDGGLYTCVATNSYGDSEQFKVNYTKPYIDETIDQTPKAVAGGEIILKCPVLGNP 3418  
QY 384 ---V---H---DL---Q---A---H-NKE-I 394  
Db 3419 TPTVTWKRGDVAENDSRHTIVANNYDLKINSVTTEDAGQVSCIADVNEAGNLTHYAAEVI 3478  
QY 395 ---Y---TI-K---W---SP--- 401  
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSPRLPSISWFRGDKPVLYDRYSISPDGSHIT 3538  
QY 402 ---T--- 402  
Db 3539 INKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIIGNPLAIVARTIYIECPI 3598  
QY 403 -G-P---G---T---NN--- 408  
Db 3599 SGIPQPDVITWTKGMDINMTDSRVILAQNNETFGIENVQVTDQGRVTCATNRGGRKASHD 3658  
QY 409 ---P--- 409  
Db 3659 FSLDVLSPPEFDIHGTOPTIKREGDTITLTCPKLAEDIADQVMDVSWTKDSRALDGLT 3718  
QY 410 -N---A---N---L---ML-A-SAS-F---D--- 421  
Db 3719 DNVDISDDGKLTISQASLENAGLYTCIALNRAGEASLEKVELSPPPVIDISRNDVQOQ 3778  
QY 422 ---S---T---VR---L---W--- 427  
Db 3779 VAVNQPTIMRCAVTHGHPFSIKWLKNGKEVTDENIRIVEQGOVLQILRTDSHAGKWSC 3838  
QY 428 ---DV---D---R-G-I-C--- 434  
Db 3839 VAENDAGVKEMLVDVFTPPVSVKSDNPRIKALGETITLFCNAGSNPPQLKWAAGGSL 3898  
QY 435 ---I-H---T--- 437  
Db 3899 IFDSPDGRISLKGARLDIPLHKKTDVGYTCOALNAAGTSEASVSVLVLPPEINRDGI 3958  
QY 438 ---LT---K---H---Q--- 442  
Db 3959 DMSPLPAQOGLTLQCLAQKQPVQMRWTLNGTALTSTPTITVADSTFIQINNVSLSD 4018  
QY 443 ---E---PV---Y---S--- 447  
Db 4019 KGYVTCYAENVAGSDNLMNVVQAPVISNGGTKQVIEGELAVIECLVEGYPAPQVSWL 4078

QY 448 ---Y---A--- 449  
Db 4079 RUGNRVETGVQGVYVYTDGRMLTIIIEARSLDSGILCSATNEAGSAQAQVTLVLSPKI 4138  
QY 450 ---PS---P---DG--- 454  
Db 4139 ITSTPGVLTPSSGSKFSLPCAARGYPDPPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4198  
QY 455 -R-Y---L---A---SG--- 460  
Db 4199 ERHLIYECTAKNDAGADTLEFPVQTVIAPKISTISGNRYINGSEGTETVIKCEIESSESSEF 4258  
QY 461 ---S---F-D-K---C--- 465  
Db 4259 SMSKNGVPLLPNNLIFSEDYKLIKILSTRLSQGEYSCTAANKAGNATKTNLNVGVAP 4318  
QY 466 ---VH---I-W---NT--- 471  
Db 4319 KIMERPRTOVHKGDQVTLWCEASGVQPAITWYKDNELLTNTGVDDETATTKKSVIFSS 4378  
QY 472 ---Q--- 474  
Db 4379 ISPSQAGVYTKAENVVASTEEDIDLIVMIPPEVPERMNVSTNPRQTVPLSCNATGIPE 4438  
QY 475 ---L---V---H--- 478  
Db 4439 PVISWMRDSNIAIQNNEKYQILGTLTIAIRNVLPDDGDFYHCIAKSDAGOKIATKRLIVNK 4498  
QY 479 ---S-YR---G- 482  
Db 4499 PSDRPAPIWECEDEKGPKEKTEYMDRGDTPDDNPQLPEKWDVEDSSLNGSIAYRCMPGP 4558  
QY 483 ---T---G---G-I---FE--- 488  
Db 4559 RSRTVLLHAAPQFIVKPKNTTAAIGAIVELRCSAAGPPHTTITWAKDGKLIEDSKFEIA 4618  
QY 489 ---V-C--- 490  
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSAPINVDNNILPTPKSSNQKNVAVITCYE 4678  
QY 491 ---W---N-A--- 493  
Db 4679 RNQAVSRGLTWEYNGVPMKPNLAGIHFMMNGSILVILDTSSLKEGDLTYTCKVRNRRHS 4738  
QY 494 ---A-G---DKV---G---AS--- 502  
Db 4739 IPHLTSAFEGPEVKTIDKVENNGSVVLDCEVTSDDLTHVVTWKNQKMLDDDAIYV 4798  
QY 503 ---S---D-GSV-CV---LD---L---R---K 514  
Db 4799 LPNNSILVLLNVEKYDEG-VYKCVASNSIGKAFDDTQNLNVYGGSSRREAYK 4847

## RESULT 14

US-10-120-801-75

; Sequence 75, Application US/10120801

; Publication No. US20030203843A1

; GENERAL INFORMATION:

; APPLICANT: Pena, Carol

; APPLICANT: Guo, Xiaojia

; APPLICANT: Shimkets, Richard

; APPLICANT: Padigar, Muralidhara

; APPLICANT: Kekuda, Ramesh

; APPLICANT: Spytek, Kimberly

; APPLICANT: Mehraban, Fuad

; APPLICANT: Topper, James N.

; APPLICANT: Malyankar, Uriel

; APPLICANT: Wasserman, Scott

; APPLICANT: Edinger, Shlomit

; APPLICANT: Smithson, Glenn

; APPLICANT: Gunther, Erik

; APPLICANT: Komuves, Laszlo

; TITLE OF INVENTION: Proteins and Nucleic Acids Encoding Same

FILE REFERENCE: 21402-340  
CURRENT APPLICATION NUMBER: US/10/120,801  
CURRENT FILING DATE: 2002-04-11  
PRIOR APPLICATION NUMBER: 60/285748  
PRIOR FILING DATE: 2001-04-23  
PRIOR APPLICATION NUMBER: 60/286068  
PRIOR FILING DATE: 2001-04-24  
PRIOR APPLICATION NUMBER: 60/286292  
PRIOR FILING DATE: 2001-04-25  
PRIOR APPLICATION NUMBER: 60/288334  
PRIOR FILING DATE: 2001-05-03  
PRIOR APPLICATION NUMBER: 60/291241  
PRIOR FILING DATE: 2001-05-16  
PRIOR APPLICATION NUMBER: 60/322284  
PRIOR FILING DATE: 2001-09-14  
PRIOR APPLICATION NUMBER: 60/285609  
PRIOR FILING DATE: 2001-04-20  
NUMBER OF SEQ ID NOS: 155  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 75  
LENGTH: 5198  
TYPE: PRT  
ORGANISM: Caenorhabditis elegans  
US-10-120-801-75

Query Match 71.2%; Score 2610.8; DB 14; Length 5198;  
Best Local Similarity 9.2%; Pred. No. 1.3e-29;  
Matches 447; Conservative 52; Mismatches 12; Indels 4362; Gaps 387;

QY 1 M-----SI-----SS-----D-----E-----7  
DB 1 MGRSPSWLYGVGLLLATTCCSSVNDKNDPTKSSLAFVFDITGSMFDDLVQVREGAAK 60  
QY 8 V-----N-----F-----L-----V-----13  
DB 61 IFKTVMAOREKLIYIMVPHDPYLGEIINTTSTYFMRQLSKVYVHGGGDCPEKLTG 120  
QY 14 -----R-----Y-----L-----O-----S-----F-----SHSAF-T 27  
DB 121 ILKALQISLPSSFIYVFTDARKDKYHLEDEVLNTIQEKQSSVVFVMTGCGNTHPGFRT 180  
QY 28 -----FG-I-----E-----S-----H-----ISQ-----S-----NI-----39  
DB 181 YEKIAAASFGQVHLEKSDVSTVLEYVRHAVKQKVHLMYEARERGTVSRNIPVDKHL 240  
QY 40 -----N-----N-----GA-----LV-----P-----P-----46  
DB 241 ELTISLGDKDSNLDIVLRDPEGRVTDKRLYSKEGGTIDLKNVKLIRLKDPSGVWTV 300  
QY 47 -----A-----A-----A-----LI-----S-----I-----52  
DB 301 NTHSRLKHTIRVPGHNAVDFKYGFSRPLDRIELARPRVPLNODTYLLINMTGLIPPGTV 360  
QY 53 --I-----Q-----KGL-----Q-----Y-----VE-----AE-----63  
DB 361 GEIDLVDYHCHSLYKAVASPHRTNPNMYFAGFPVPGKFLFVRVQGYDQEDNYEFMRIA 420  
QY 64 -----V-----S-----I-----N-----ED-----GTLF-----73  
DB 421 AIGSVIVGGPRAFMSPHQBFVGRDLNLSCTVESASAYTIYVWKTGEDIIGGFLVHNTD 480  
QY 74 -----D-----D-----G-----R-----P-----I-----E-----79  
DB 481 TSVMTIPELSLKAGEVECEVISNNGNYSVKTRVETRESPEIFGVRNVSVPLGEAFLH 540  
QY 80 -----S-----L-----I-----DA-----86  
DB 541 CSTRSAGEVEIRTRYGATVNGENTERNPTNGTLKLIHVHTRADAGVYECMARNAGMST 600  
QY 87 -----VM-----P-----DV-----V-----92  
DB 601 RKMRLDIMEPPSVKVTQDQVYFNNREGVNLSCAMGDPKPEVHWYFKGRHLNDYKYQVG 660

QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----Q-----105  
DB 661 QDSKFLYIRDAATHDEGTVECRAMSQAGQA-RDITDLMLATPPKVEIILQNMWVGRGDRV 719  
QY 106 -----Q-----H-----A-----AAA-----AA-----113  
DB 720 SPECTIRGKPHPKIRWFKNGKDLKPDYIIKINEQOLHMGAKDBADAGAYSCVGNMAG 779  
QY 114 --AAAA-----T-----N-----Q-----Q-----GS-----A-----KN-----126  
DB 780 KDVQVANLSVGRVPTIIESPHTRVNIERQVTLQCLAVGIPPEIEWQKGNVLLATLNP 839  
QY 127 -----GEN-----T-----A-----N-----132  
DB 840 RYTQLADG-NLLITDAQIEDQOFTCIARNTYQQSQSSTLMTVLVSPVLGHVPPEQL 898  
QY 133 -----G-----G-----E-----N-----G-----137  
DB 899 IEGQDLTSCVVVLGTPKPSIVMIKDDKPVBEPTIKIEGGGSLRLRGNPKDEKGYTC 958  
QY 138 -----A-----H-----T-----I-----A-----N-----NH-----145  
DB 959 IAVSPAGNSTLHINVOLIKKPEFVYKPEGIVFKPTISGMDKXHAVNSTHVDLGE 1018  
QY 146 -----T-----146  
DB 1019 AIPCVSSTGTPPIITWYLDGRPIITNSRDTVTADNTLIVRKADKSYSGVYTCQATNSAG 1078  
QY 147 D-----M-----M-----E-----VD-----GD-----V-----155  
DB 1079 DNEQKTTIRIMTPIISPGQSSFNMYVDDLFTPCDVYGDPKPVITWLDDKPFTEGVN 1138  
QY 156 E-----I-----PS-----N-----K-----A-----VVL-----165  
DB 1139 EGSLLTIPVNEAHRGTFTHAQAAGNDTRVTLVHTTPTINAENQEKIALQNDIVL 1198  
QY 166 -----R-----167  
DB 1199 ECPAKALPPPVLWTYEGEKIDSLIPHTIREDGALVQNVKLENTGVFVQCVSNLAGE 1258  
QY 168 -----HE-----SEV-----F-----I-----C-----A-----WNP-----V-----180  
DB 1259 SLSYTLTVHEKPIISEVPGVVDWVGKFTIEIPCRATGVPEVIRTNKNGIDLKMDEK 1318  
QY 181 S-D-L-L-V-S-G-----SG-----D-----S-----TA-----193  
DB 1319 SVDNLGTLRIYEADKNDIGNYCVTVNEAGTSQMTTHVDVQEPPIILPSTQTNNAVVD 1378  
QY 194 RI-----W-----NL-----S-----ENST-----S-----204  
DB 1379 RVELKCYEASPPASVTFWFRGIAIGTDKGYVVEDSGTLVIQASVEDATIVTCKASNP 1438  
QY 205 -G-----P-----TQ-----L-----V-----L-----211  
DB 1439 AGKAEANLQVTVIASPIKDPDVVTVQESIKESHPSFLYCPVFSNPLFQISWYLNKPLID 1498  
QY 212 -----R-----H-----CI-----R-----216  
DB 1499 DKTSWKTSDDKRLHVFKAKITDSGVYKCVARNAGEGSKSFQEVIVPLNLDESXYKK 1558  
QY 217 -----EG-----G-----Q-----D-----V-----P-----SN-----225  
DB 1559 VPAKEGEVTLGCPVSGFPVQPINWVVDGTVVBPVKYKATLSNGLTLHDFSVSVKQE 1618  
QY 226 -----KDVIT-S-L-----D-----W-----233  
DB 1619 GNYHCAQSKGNILIDIVELSLAVPIVGEDDNLVFLGKDISLSDLOTESDCKTTFW 1678  
QY 234 --N-SB-----G-----TL-----L-----240  
DB 1679 SINGSERPDNVQIIPSDGHRLYITDAKPENNGKYMCRVNTNSAGKAERTLTLDVLEPPV 1738  
QY 241 -----A-----TG-----SY-----D-----246

Db 1739 VEPVFANQKLGNNPIILOQVGTGNPKPTVIWKIDGNDVDKSWLFDESLSLRIBKLTG 1798  
QY 247 -G- -F- -A- -R- -I- 251  
Db 1799 KSAQISCTAENKAGTASRDFQIQAAPTAKNEGQDQETIPRESEITLDCPVSGLGFOIT 1958  
QY 252 W- -T- -KD- -S- 255  
Db 1859 WMKQGLPLTENDAIPTLNDTRLTLNANRDHDIYTCVANNTAGQVSKDFVVQVLPKI 1918  
QY 256 -GN- -L- -A- -S- 260  
Db 1919 KNAVVTLEINEGEIILTCAENPTPTAKWDFNQDGLPKAEAFVNNHTVVVVNNVTKYH 1978  
QY 261 T- -L- -GO- -H- -K- - 266  
Db 1979 TGUYKYATNKVGQAVKTINNVHRTKPRFESGLTESELTVNLTRSLTECDVDDAIGVGI 2038  
QY 267 -G- -P- -I- -F- 270  
Db 2039 SWTNGKPFLEAETDGVOTLAGRFLHIVSAKTDHGSYACTVTNEAGVATKTNLFQVVP 2098  
QY 271 -AL- -K- -W- - 274  
Db 2099 PTIVNEGGEYVTIENNSLVLPCBVTGKPNPVVWTGDRPVGLKSVQVLSGQOQFKIVH 2158  
QY 275 -N- -K- -KG- - 278  
Db 2159 AEIAHKSYICMAKNDVGTABISFVDIITPRMIQKINIVTAIRGGALPFKCPIDDDK 2218  
QY 279 NF- -IL- -S- -AG- 285  
Db 2219 NFKGQIILWNYQPIDLEADARITLSNDRRLTILNVTENDEGQVSCRVDNAGENSFD 2278  
QY 286 -V- -D- -KT- -TI- -I- -W- -D- -A- -H- - 296  
Db 2279 FKATVLVPPTIIMLDKDKNTAVEHSTVTLSCPATGKPEPDITWFKDGEAIHENTADII 2338  
QY 297 -T- -G- -EA- -K- -Q 302  
Db 2339 PNGELNGNQLKITRIKEGDAGKYTCADNSAGSVEQDVNVNVTIPKIEKDGPISDYESQ 2398  
QY 303 Q- -F- - 304  
Db 2399 QNERVVISCPVYAPPAKITWLAGRPLQSDKFKVKTSAHQKLYLKLRETSSSKYTCTIA 2458  
QY 305 -P- -F- -H- -S- 308  
Db 2459 TNEAGTDKDRDFKVSMLVAFSDFEPNIVRRITVNSGNPSTLHCPAKGSPSTITWLDKDNA 2518  
QY 309 -A- -P- - 310  
Db 2519 IEPNDRYVFDAGRLQISKTEGDOGRVTCIATNSVSGDDLENTLEVIIPVIDGERRE 2578  
QY 311 -A- -L- -D- -VD- -WO- -S- -N- - 319  
Db 2579 AVAVIEGFSSELCDSNSTGVVWQKQDGLTINQDTRLGDSFIQIPSSGKKMSFLSARKS 2638  
QY 320 -N- -T- -FA- -S- -C- -S- -T- - 327  
Db 2639 DSGRYTCIVRNPAEARKLFDFAVNDPPSISDELSSANIQTIIVPYVPVEINCVVSGSPHP 2698  
QY 328 -DM- -C- -I- 331  
Db 2699 KVTWLFDDKPLEPDSAAYELTNNGETLKIVRQVEHAGVTYCEAONNVGKARKDFLVRVT 2758  
QY 332 -H- -V- -C- -KL- -G- 337  
Db 2759 APHFKEREEVVARVGDITMLTLCNAESSVPLSSVYVHHADESQVNGVITSKYAANEKTL 2818  
QY 338 -Q- - - -DR- -PI- -K- - 343

Db 2819 NVTNIQLDDGEFYCTAVNAGITKFKFLIVIETPYFLDQKQYPIILGKRLLTDCSAT 2878  
QY 344 -T- -F- - 345  
Db 2879 GTPPPITLFWKDGKRLNESDEVDIIIGSTLVDNPNQKEVEGRYTICIAENKAGRSEKMMVE 2938  
QY 346 -Q- -G- -H- - 348  
Db 2939 VLLPPKLSKEWINVEYQAGDPLTECPIDBTSGVHITWSRQFGKQDGLDMRAQSSSDSKS 2998  
QY 349 -T- -N- -E- -VN- -A- 354  
Db 2999 LYIMOATPEDADSYSCIAVNDAGBAEFQVTVNTPPKIFGDSFSTTEIVADTTLEIPCR 3058  
QY 355 -IKW- - 358  
Db 3059 TEGIPPEISLWFLDGKPILEMPGVYTKQGDLSLRIDNIKNQEGRYTCVAENKAGRAEQD 3118  
QY 359 -P- -T- -GN- - 362  
Db 3119 TYVEISEPPRVVMASEVMRVVEGRQTTIRCEVFGNPEPVVNMWKDGEPTYDLSLQFSTKL 3178  
QY 363 -L- -LA- -SC- -S- -D- -D- -M- 371  
Db 3179 SYLHLRETTLADGGTYTTCIATNKAGESQTTTQVEVLVPPRIEDEERVLOGKEGNTYMHVC 3238  
QY 372 -T- -L- - 373  
Db 3239 QVTGRPVVYVTKRNGKBEIQPNVLHNRATRADEGKYSCIASNAGTAVADFLIDVFT 3298  
QY 374 -KI- -W- -S- -M- -K 379  
Db 3299 KPTFETHETTFNIVEGESAKIECKIDGHPKPTISWLKGGPFNMNDIILSPRGDTLMILK 3358  
QY 380 -QD- -N- -C- - 383  
Db 3359 AQRFDGGLYTCVATNSYGDSEQDFKNVYTKPYIDETIDQTPKAVAGEIILKCPVLGNP 3418  
QY 384 -V- -H- -DL- -O- -A- -H- -NKE- -I 394  
Db 3419 TPTVTWKRGGDVAVPNDSRHTIYNNYDLKINSVTTEDAGYSCIAVNEAGNLTHYAAEVI 3478  
QY 395 -Y- -TI- -K- -W- -SP- - 401  
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKRPVLYDRYSISPDGSHIT 3538  
QY 402 -T- - 402  
Db 3539 INKAKLSDGGKYICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIYLECPI 3598  
QY 403 -G- -P- -T- -NN- - 408  
Db 3599 SGIPQPDVITWTKNGMDINMTDSRVILAQNNTFGIENVQVTDQGRYTCTATNRGKASHD 3658  
QY 409 -P- - 409  
Db 3659 FSLDLVSPPEFDHGTQPTIKREGDTITLTCPILAEADIAQVMDVSVWTKDSRALDGLT 3718  
QY 410 -N- -A- -N- -L- -ML- -A- -SAS- -F- -D- - 421  
Db 3719 DNVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPPVIDISRNDVQPO 3778  
QY 422 -S- -T- -VR- -L- -W- - 427  
Db 3779 VAVNOPTIMRCVGTGHPFPSIKWLKNGKEVTDDENIRIVEQGVQLQILRTSDSHAGKWSC 3838  
QY 428 -DV- -D- -R- -G- -I- -C- - 434  
Db 3839 VAENDAGVKELEMLVDVFTPPVSVKSDNPRIKALGETITLFCNASGNPYQLKWAAGGSL 3898  
QY 435 -I- -H- -T- - 437  
Db 3899 IFDSPDGAISLKGARLDIPHLKKTVDGVTCTQALNAGTSEASVSDVLPVPEINRDGI 3958

QY 438 -----LT-----K-----H-----Q----- 442  
Db 3959 DMSPLPQAQSLTLQCLAQKPPQMRWTLNGTALHTSPGTVASDSFIQINNVLSLD 4018  
QY 443 -----E-----PV-----Y-----S----- 447  
Db 4019 KGYYTCAENVAGSDNLNMYVVDVQAPVISNGGKQVIEGELAVIECLVEGYPAPQVSWL 4078  
QY 448 -----V-----A----- 449  
Db 4079 RGNRVETGVGVYVYDGRMLTIIEARSLDSGIYLCSATNEAGSAQAYTLEVLSPKI 4138  
QY 450 -----FS-----P-----DG----- 454  
Db 4139 ITSTPGVLPSSGSKSLPCAARGVDPPIISWTLNGNDIKDGENGTIGADGTLHIEKAE 4198  
QY 455 -R-Y-----L-----A-----SG----- 460  
Db 4199 ERHLIVECTAKNDAGADTLEFPVQTVIVAPKISTSGNRYINGSEGTETVIKCEIESESSEF 4258  
QY 461 -----S-----F-----D-----K-----C----- 465  
Db 4259 SMSKNGVPLPSNNLIFSEDYKLIKILSTRLSDQGEYSCTAANKAGNATQKTNLVGVAP 4318  
QY 466 -----VH-----I-----W-----NT----- 471  
Db 4319 KIMERPTQVHHKQDQVTLWCEASGVPOPAITWYKDNELLTNGVDETATTKKSVIFSS 4378  
QY 472 -----A-----L-----V-----H----- 478  
Db 4379 IPSQAGVYTCKAENWVASTEEDILIVMIPPEVPERMVNSTPQTVFLSCNATGIPE 4438  
QY 475 -----A-----L-----V-----H----- 478  
Db 4439 PVISWMRDSNIAIQNEKYQILGTTLAIRNVLPPDDGFGYHCIAKSADAGQKIATRKLIIVNK 4498  
QY 479 -----A-----T-----G-----G-----I-----FE----- 488  
Db 4499 PSRPAPIWVECEDEKPKKTEYMWIDRGDTPDDNPQLLPKQVDESSLNGSIAYRCMPGP 4558  
QY 483 -----A-----G-----G-----G-----I-----FE----- 488  
Db 4559 RSSRTVLLHAAPQFIVKPKNTTAAIGAIVELRCSAAGPPHPTITWAKDGKLIBDSKFEIA 4618  
QY 489 -----W-----N-----A----- 493  
Db 4619 YSHLKVTLNSTSDSEYTCMAQNSVGSSTVSFAFVNDNNILPTPKPSSNKNVAVITCYE 4678  
QY 491 -----W-----N-----A----- 493  
Db 4679 RNQAYSRLTWYNGVMPKXNLAGIHPMNGSLVILDTSSLKEGDLELYTCKVNRNRHS 4738  
QY 494 -----A-----G-----DKV-----G-----AS----- 502  
Db 4739 IPHLTSAFEGVPEKTIKQVNVNGDSVWLDCEVTSDELTHVVTWKQKMLDDDAIYV 4798  
QY 503 -----S-----D-----GSV-----CV-----L-----D-----L----- 512  
Db 4799 LPNNSVLLNVEKYDEG-VYKCVASNSIGKAFDDTQLNVYEGDFLPTGFEFGSGINIDS 4857  
QY 513 -----R-----K----- 514  
Db 4858 SNAGSSRRREAYK 4870

## RESULT 15

US-10-369-493-6858  
; Sequence 6858, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; TITLE OF INVENTION: 38-10(52052)B  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; PRIOR FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 6858  
; LENGTH: 5198  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
; US-10-369-493-6858

Query Match 71.2%; Score 2610.8; DB 14; Length 5198;  
Best Local Similarity 9.2%; Pred. No. 1.3e-29;  
Matches 447; Conservative 52; Mismatches 12; Indels 4362; Gaps 387;  
QY 1 M-----SI-----SS-----D-----E----- 7  
Db 1 MGRSPSLVGLVGLLLATTCSVNDKNDPTGKSLAFVFDITGSMFDDLVQREGAAK 60  
QY 8 ---V-----N-----F-----L-----VY----- 13  
Db 61 IPKTVMAQREKLIYNYVPHDPYLGELIINTDSTYFMRQLSKVYVHGGDCPEKTLTG 120  
QY 14 ---Y-----R-----Y-----L-----OE-----SG-----F-----SHSAP-T 27  
Db 121 ILKALQISLPSSFYVFTDARKDYHLEDEVLTIOEKQSVVFMVMTGCGNRTHPGFRT 180  
QY 28 ---PG-I---E---S---H-ISO-----S-NI----- 39  
Db 181 YEKIAAASFGVPHLEKSDVSTVLEYRHAVKQKVHLMYERARCGTVSRNIPVDKHL 240  
QY 40 ---N-----N-----GA-----LV-----P-P----- 46  
Db 241 ELTISLKGKDDSDNLDIVLRDPEGRTVKRLYKSGGTIDLNKVKILRLKDPSPGVTV 300  
QY 47 ---A-----A-----A-----LI-----S-I----- 52  
Db 301 NTNSRLKHTIRVFGHGAVDKYGAFASRLDRIELARFVLNQDTYLLINMTGLIPPGTV 360  
QY 53 ---I-----Q-----KGL-----Q-Y-----VB-----AE- 63  
Db 361 GEIDLVDYHSHLYKAVASPHRTNPNMYFAGFPVPPKGLFFVTVQGYDEDNYEFMRIA 420  
QY 64 ---V-----S-I-----N-----ED-----GTLF----- 73  
Db 421 AIGSVIVGGPRAFPMSPIHQEFVGRDLNLSCTVESASAYTIYWKGTEDIIGGFLFYHNTD 480  
QY 74 ---D-----D-----G-----R-----P-I-----E----- 79  
Db 481 TSVMTIPELSLKDAGEYECRVISNNNGSVKTRVETRESPEIFGVNRVNSVPLGEAFLH 540  
QY 80 ---S-----L-----SL-I-----DA----- 86  
Db 541 CSTRSAGEVEIRWTRGVAFVNGPNTERNPTNGTLKIHVTRADAGVYECMARNAGMST 600  
QY 87 ---VM-----P-DV----- 92  
Db 601 RKMRLDIMEPPSVKVTTPQDYFNMREGVNLSCAMGDKPEVHWYFKGRHLLNDYKQVG 660  
QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----Q----- 105  
Db 661 QDSKFLYIRDATHHDEGTVECRAMSQAQA-RDITDMLATPPKVEIIONKMMVGRGDRV 719  
QY 106 ---S-----Q-H-----A-----AAA-----AA- 113  
Db 720 SFEKTRKPKPKIRWFKNGKDLIKPDDYIKINEGQLHIMGAKDEBAGAYSCUGENMAG 779



114 QY --AAAA--T--N--Q--Q--GS--A--KN--126  
780 DB KDQVANLSVGRVPTTIESPTVRVNIERQVTLQCLAVGIPPEIEBQXGNVLLATLNNP 839  
127 QY --GEN--T--A--N--132  
840 DB RYTQLADG-NLLITDAQIEDQOFTCIARNTYQQSQSTTLMVTLGLVSPVGHVPPPEQL 898  
133 QY --G--G--EE--N--G--137  
899 DB IEQODLTSCVVVLGTPKPSIWKDKDPVEEGPTIKIEGGGSLLLRGNPKDEKGYTC 958  
138 QY --A--H--TI--A--N--NH--145  
959 DB IAVSPAGNSTLHINVQLIKKPEFVYKPEGGIVFKPTISGMDEKHVAVNSTHVDLDEGF 1018  
146 QY --G--F--T--146  
1019 DB AIPCVCVSGTPPIITWYLDGRPTPNRSDFTVTADNTLIVRKADKSYSGVYTQATNSAG 1078  
147 QY --M--M--E--VD--GD--V--155  
1079 DB DNEQKTIIRMTPMISPGOSSFNWVDDLFTIPCDVYGDPKVITWLLDDKPFTEGVN 1138  
156 QY --I--PS--N--K--A--VVL--165  
1139 DB EDGSLTIPVNEAHRGTFTCHAQNAAGNDRTRVTLVHTTPTTINAENQEKIALQNDIVL 1198  
166 QY --R--G--167  
1199 DB ECPAKALPPPRLWYTEGEKIDSQIPHTIREDGALVLQNVKLENTGVFCVQVSNLAGED 1258  
168 QY --HE--SEV--F--I--C--A--WNP--V--180  
1259 DB SUSYTLTVHEKPIISEVPGVDVVGKFTIEIPCRATGVPEVIRTWKNGIDILKMDKKP 1318  
181 QY S-D-L--L--V--S--G--D--S--TA--193  
1319 DB SVDNLGLTIRIYEADKNDIGNYCVVTVNEAGTSQMTTHVDVQBPPIILPSTQTNNTAVGD 1378  
194 QY RI--W--NL--S--ENST--S--204  
1379 DB RVELKCYVEASPPASVTFWRRGIAIGTDTKGYVVEDGTLVIOQASVEDATIYTCASNP 1438  
205 QY --G--P--TO--L--V--L--211  
1439 DB AGKAEANLQVTVIASPDIKDPDVVTOESIKESHPSLYCPVFSNPULPOISWYLNKPLID 1498  
212 QY --R--H--CI--R--216  
1499 DB DKTSMKTSDDKRKLHVFKAKITDSGVYKCVARNAAGEGSKSFOVEIVPLNLDESXYKKK 1558  
217 QY --EG--G--Q--D--V--P--SN--225  
1559 DB VFAKEGEVTLGCPVSGFPVQINWVDTGVVBPBKGYKATLSNDGLTLHFDVSXVQOE 1618  
226 QY --K--S--L--D--W--233  
1619 DB GNVHCVAQSKGNLIDIVLSVLAVPIVGEDDNLVFLGKDISLSDLOTESDDKTTFFW 1678  
234 QY --N--SE--G--TL--L--240  
1679 DB SINGSESDRPNVQIPSDGHLRYITDAKPNNGKYMCRVNSAGKAERTILTLVDLEPPVF 1738  
241 QY --A--TG--SY--D--246  
1739 DB VEPVFEANQKLIGNNPILQCVQTNPKPTVIWKIDGNDVDKSWLFDLSLLRIEKLATG 1798  
247 QY --G--F--A--R--I--251  
1799 DB KSAQISCTAENKAGTASRDFFIQNIAPTFKNEGQDQETIFRESEITLPCVSLGDFQIT 1858  
252 QY W--T--KO--255

1859 DB WMKQGLPLTENDAIFTLDNTRLTILNANRDHEDIYTCVANNTAGOVSKDFVVVVQVLPKI 1918  
256 QY --GN--L--A--S--260  
1919 DB KNNAVTLNEBEEIILCTDAEIGNPPTAKWDFNOGDLKPEAVFVNNHTVVVNNVTKYH 1978  
261 QY --L--Q--H--K--266  
1979 DB TGVYKCYATNKVGQAVKTNVHVTKPFESGLTSELTNLNLTSTRSITLBCDVDDAIGVGI 2038  
267 QY --G--P--I--F--270  
2039 DB SWTVNGKPLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVTNEAGVATKTNLNFVQVP 2098  
271 QY --AL--K--W--274  
2099 DB PTIVNEGGEYTVIENNSLVLPCEVTGKPNPVVTWTKGRPVGDLKSVQVLSGQQPKIVH 2158  
275 QY --N--K--KG--278  
2159 DB AEIAHKGSYICAKNDVGTAEISFDVDIITRPMIOKGIKNIIVTAIKGGALPKFKPIDDDK 2218  
279 QY NF--IL--S--AG--285  
2219 DB NFKGQIILWRNYQPIDLEADARITRLSNDRLTILNVNTEDEGOYSCRVKNDAGENSFD 2278  
286 QY --V--D--KT--TI--I--W--D--A--H--296  
2279 DB FKATVLVPPTIIMLDKDNKTAVEHSTVTLSCPATGKPEPDITWFKDGEAIIHENIADII 2338  
297 QY --T--G--EA--K--Q--302  
2339 DB PNGELNGNLKLTIRIKEGDAGKYTCADNSAGSVEQDVNVNVTIIPKIEKOGIPSDYESQ 2398  
303 QY --P--P--H--S--308  
2459 DB TNEAGTKRDFKVSMLVAPSPDEPNIVRITVNSGNPSTLHCPAKGSPPTITWLDKQNA 2518  
309 QY --A--P--310  
2519 DB IEPNDRYVFFDAGRLQISKTEGSDQGRYTCIATNSVGSDDLLENTLEVIIPPIDGERRE 2578  
311 QY --A--L--D--VD--WQ--S--N--319  
2579 DB AVAVIEGFSSELFCDSNSTGVDVWOKGLTINQDTRGDSFIQIPSSGKKMSFLSARKS 2638  
320 QY --N--T--FA--S--C--S--T--327  
2639 DB DSGRYTCIVRNPAGEARKLFDFAVNDPPSISDELSSANIQTIVPYVPVINCVSUSGPH 2698  
328 QY --DM--C--I--331  
2699 DB KYTWLFDDKPLPDSAAVELTNGGETLKIVRSQVEHAGTYTCEAQNNGVKARKDFLVRVT 2758  
332 QY --H--V--C--KL--G--337  
2759 DB APPHFEKEREVVARVGDMLTLCNAESSVPLSSVYVWHAHDESQVGVITSKYAANEKTL 2818  
338 QY --Q--DR--PI--K--343  
2819 DB NVTNIQLDDEGFYCYTAVNEAGITKFKPLIVIETPYFLDQOKLYPIILGKRLTDCSAT 2878  
344 QY --T--P--345  
2879 DB GTFPPPTILFMKDKRLNDESDVDIIGSTLVIDNPQVEGRTYCIENKAGREKMDMVE 2938  
346 QY --Q--G--H--348

Db 2939 VLLPPLSKWENVEQADPLTLECPIDTSGVHITWSRQFGKQGLDMRAQSSDKSK 2998  
QY 349 -----T-----N-----E-----VN-----A----- 354  
Db 2999 LYINQATPEDADSYCIAVNDAGAEAVQVNTNPPKIFGDSFSTTEIVADTLEIPCR 3058  
QY 355 -----IKW-----P-----S-----P-----DG----- 454  
Db 3059 TEGIPPEISWFLDGKPILEMPGYTKQGLSLRIDNIKPNQEGRYTCVAENKAGRAEQD 3118  
QY 359 -----P-----T-----GN----- 362  
Db 3119 TYEISPPRVWMASEVMRVVEGRQTTIRCEVFENPVPVNWLDKGBPYTDLQPSKTL 3178  
QY 363 -----LA-----SC-----S-----D-----M----- 371  
Db 3179 SYLHRETTLAGDGYTCIATNKAGESOTTDEVLVPPRIEDBEERVLOGKEGNTYMHVC 3238  
QY 372 -----L-----T-----L----- 373  
Db 3239 QVTCRPVYVTKENGKEIQFNPVLHNRATRADEKYSCIASNEAGTAVADFLDVFT 3298  
QY 374 -----V-----H-----DL-----Q-----A-----H-NKE-I 394  
Db 3419 TPTVTWKRGDVAPNDSRHITVNNYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVI 3478  
QY 395 -----Y-----TI-K-----W-----SP----- 401  
Db 3479 GKPTFVRKGNLYEVIENDTITMDCGVTSRPLPSISWFRGDKPVLYDRYSISPDGSHIT 3538  
QY 402 -----V-----H-----DL-----Q-----A-----H-NKE-I 394  
Db 3539 INKALSDGGKYICRASNEAGTSDIDILKILVPPKIDKNSIIGNPLAIVARTIYLECPI 3598  
QY 403 -----G-----T-----NN----- 408  
Db 3599 SGIPQPDVITWKGMDINMTDSRVILAQNNETFGIENVQVTDQGRYCTATNREGKASHD 3658  
QY 409 -----P----- 409  
Db 3659 FSLDVLSPFPDIHGTOPTIKREGDTITLTCPIKLAEDIAQVMDVSWTKDSRALDGLT 3718  
QY 410 -----N-----A-----N-L-----ML-----A-SAS-----F-----D----- 421  
Db 3719 DNVDISDGRKLTISQASLENAGLYTCIALNRAGEASLEPKVILSPFPVIDISRNDVQPQ 3778  
QY 422 -----S-----T-----VR-----L-----W----- 427  
Db 3779 VAVNQPTIMECAVTHGHPFPIKMLKNGKEVTDENIRIVEQGVQLILRTDSHDAGKWSC 3838  
QY 428 -----DV-----D-----R-----G-----I-----C----- 434  
Db 3839 VAENDAGVKEMLVDVFTFPVSVKSDNPICALGETITLFCNASGNPYQLKWKAGGSL 3898  
QY 435 -----I-H-----T----- 437  
Db 3899 IFDSPGARIKLGARLDIPLHKKTVDGYTCQALNAGTSEASVSDVLVPPPEINRDI 3958  
QY 438 -----LT-----K-----H-----Q----- 442  
Db 3959 DMSPLPAQOQSLTLQCLAQCKPVPQMRWTLNGTALTHTSTPGITVASDSTFIQINNVSLS 4018  
QY 443 -----E-----PV-----Y-----S----- 447  
Db 4019 KGVYTCYAENVAGSDNLMYNDVVVQAEVVISNGGTQVIEGELAVIECLVEGYPAPQVSWL 4078

QY 448 -----V-----A----- 449  
Db 4079 RGNRVETGVQVRYVYTDGRMLTIIIEARSLDSGIYLCSATNEAGSAQAYTLEVLVSPKI 4138  
QY 450 -----PS-----P-----DG----- 454  
Db 4139 ITSTPGVLPSSGKESLPCAVRGYPDPPIISWTLNGNDIKDGENGHTIGADGTLHIEKAE 4198  
QY 455 -----R-----Y-----L-----A-----SG----- 460  
Db 4199 ERHLIYEKAKNDAGADTLEFPVQTTIVAPKISTSGNRYNGSEGTETVIKCEISESESEF 4258  
QY 461 -----S-----F-----D-----K-----C----- 465  
Db 4259 SMSKNGVPLLPNNLIPFSEDIKLIKILSTRLSQGBYSCTAANKAGNATQKTNLNVGAP 4318  
QY 466 -----VH-----I-W-----NT----- 471  
Db 4319 KIMERPTQVVKGDQVTLWCEASGYPQPAITWYKDNELLTNTGVDETATTTKKSVIFSS 4378  
QY 472 -----A-----L-----V-----H----- 478  
Db 4439 PVISWNRDSNIALQNNKEYQILGTTLAI RNVLPPDDGDFYHCIAKSADAGOKIATRKLI VNK 4498  
QY 479 -----W-----N-A-----S-YR-----G- 482  
Db 4499 PSDRPAPIWECDEKPKKTEYMDRGTPDDNPOLLPWKQVEDSSLANGSIAYRCMPGP 4558  
QY 483 -----T-----G-----G-----I-----FE----- 488  
Db 4559 RSSRTVLLHAAPOQIVKPKNTAAIGAIVELRCSAAGPPHPTITWAKDGKLIEDSKFEIA 4618  
QY 489 -----V-----C----- 490  
Db 4619 YSHLKVTLNSTSDSGEYTCMAQNSVGSSTVSFAFINVDNNILPTPKPSSNQKNVAVITCYE 4678  
QY 491 -----W-----N-A----- 493  
Db 4679 RNOQVSRGLTWYNGVPMKPNLAGIHFMMNGSLVILDTSSLKEGDLLEYTCVKVNRNRHS 4738  
QY 494 -----A-G-----DKV-----G-----AS-----A----- 502  
Db 4739 IPHLTSAFEVPEVKTIIDKVEVNGSDSVLLDCEVTSDPLTTHVVTNKDQKMLDDAIYV 4798  
QY 503 -----S-----D-GSV-----CV-----L-----D----- 512  
Db 4799 LPNNSLVLLNVEKYDEG-VYKCVASNSICKAFDDTQNLNVEGDFLPTGTFEGSGINIDDS 4857  
QY 513 -----R-----K 514  
Db 4858 SNAGSSRREAYK 4870

Search completed: January 3, 2005, 15:49:10  
Job time : 139 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 15:09:41 ; Search time 21.6667 Seconds  
(without alignments)  
2282.558 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 3669

Sequence: 1 MSISSDEVNFLVRYLQESG.....GDKVGASDGSVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 79:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2612.1	71.2	5175	2 T20592	hypothetical prote
2	2610.8	71.2	5138	2 T43290	hemocentin precurs
3	2603.7	71.0	7962	2 I38346	elastic titin - hu
4	2594.2	70.7	6658	2 T13931	projectin - fruit
5	2593.7	70.7	10797	2 T30192	probable peptide s
6	2592.2	70.7	26926	1 I38344	titin, cardiac mus
7	2590.9	70.6	7829	2 T15789	hypothetical prote
8	2587.7	70.5	4545	1 S25111	alpha-2-macroglobu
9	2586.1	70.5	4544	1 S02392	alpha-2-macroglobu
10	2585	70.5	6831	2 A88852	protein unc-22 (im
11	2585	70.5	6839	2 S57242	twitchin (similar
12	2585	70.5	7160	2 T27935	hypothetical prote
13	2583.7	70.4	4680	2 T42377	gp330 protein prec
14	2580.1	70.3	5825	2 T12117	polyprotein - fava
15	2579.8	70.3	3461	2 S58870	reelin precursor -
16	2579.7	70.3	5376	2 T42215	zonadhesin - mouse
17	2575.1	70.2	8243	2 T31307	type I fatty acid
18	2570.7	70.1	4543	1 A53102	alpha-2-macroglobu
19	2569.3	70.0	5232	2 A45086	HC-toxin synthetas
20	2566	69.9	6805	2 S20901	titin - rabbit (fr
21	2556.6	69.7	4753	1 A47437	LDL-receptor-relat
22	2550.8	69.5	15281	2 S41309	cyclosporin synthe
23	2548.4	69.5	4302	2 A38971	polycystic kidney:
24	2547.9	69.4	6669	2 S55024	nebulin, skeletal
25	2545	69.4	4767	2 T31345	hypothetical prote
26	2543	69.3	6642	2 T29757	protein UNC-89 - C
27	2542.1	69.3	5138	2 B96695	hypothetical prote
28	2541.3	69.3	4861	2 S71752	giant protein p619
29	2538.8	69.2	3623	2 T09456	intrinsic factor-B

30	2538.3	69.2	9376	2 T14593	syngomycin synth
31	2536.9	69.1	4836	2 T14346	herc2 protein - mo
32	2528.1	68.9	4930	2 E69679	polyketide synthet
33	2526.8	68.9	4447	2 A69679	polyketide synthas
34	2525.5	68.8	3856	2 T51174	ataxia-telangiecta
35	2523.8	68.8	5369	2 T44807	mycosubtilin synth
36	2523.5	68.8	4644	1 A38905	dynein heavy chain
37	2523.2	68.8	4572	2 S57908	hypothetical 527K
38	2522.6	68.8	4568	2 T08030	dynein beta heavy
39	2520.2	68.7	4639	2 T08030	dynein heavy chain
40	2519.9	68.7	3623	2 T08618	intrinsic factor-B
41	2519	68.7	4344	1 A53489	dynein heavy chain
42	2513.1	68.5	4464	2 D87755	protein T21E12.4 f
43	2511.7	68.5	4391	2 A38096	perlecan precursor
44	2511.1	68.4	6486	2 T31076	tyrocidine synthet
45	2508.3	68.4	4273	2 C69679	polyketide synthas

## ALIGNMENTS

### RESULT 1

T20992

hypothetical protein F15G9.4a - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T20992; T24733

R:Sulston, J.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z19355

A:Accession: T20992

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5175 <WIL>

A:Cross-references: EMBL:Q810L3; EMBL:Z47068; PIDN:CAA87335.1; GSPDB:GN00028; CESP:F15G9.4a

A:Experimental source: clone F15G9

R:Kershaw, J.

submitted to the EMBL Data Library, December 1994

A:Reference number: Z19929

A:Accession: T24733

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-5175 <W12>

A:Cross-references: EMBL:Z47070; PIDN:CAA87344.1; GSPDB:GN00028; CESP:F15G9.4a

A:Experimental source: clone T09B9

C:Genetics:

A:Gene: CESP:F15G9.4a

A:Map position: X

A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3; 1251/2; 2591/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4879/1; 4941/1; 5011/1; 5077/1

Query Match 71.2%; Score 2612.1; DB 2; Length 5175;  
Best Local Similarity 9.2%; Pred. No. 6e-53;  
Matches 446; Conservative 53; Mismatches 12; Indels 4339; Gaps 386;

QY	1	M-----SI-----SS-----D-----E-----	7
DB	1	MGRSPSWLYGLGLLLALLATTCSSVNDKNDPTGKSLAFVFDITGSMFDDLVOVREGAAK	60
QY	8	-----V-----N-----F-----L-----VY-----	13
DB	61	IPKTVNAQREKLIYVIMVFFHPDPLGEINTTDSYFMRQLSKVYVHGSGDCPEKTLTG	120
QY	14	-----R-----Y-----L-----OE-----SG-----F-----SHSAF-T	27
DB	121	ILKALQISLPSSFIYVFTDARSKDYHLEDEVLNTIQEKSSVVFVMTGDCGNRTGFGFT	180
QY	28	-----EG-I-----E-----S-----H-ISO-----S-NI-----	39
DB	181	YEKIAAASFGQPHLEKSDVSTVLEYVRHAVKQKVHLMYAEARGGTVSRNIPVDKHL	240
QY	40	-----N-----GA-----LV-----P-----P-----	46

Db	241	ELTISLGDKDDSNLDIVLRDPGRVTKRLYSKEGGTIDLKNVKLIRLKDPSGWTV	300
Qy	47	-----A-----A-----LI--S--I-----	52
Db	301	NTNSRLKHTIRVFGHNAVDPKYGFASRPDLRIELARPRVLNODTYLLINMTGLIPPGTV	360
Qy	53	-----I-----Q-----KGL-----Q-Y-----VE-----AE-----	63
Db	361	GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPVPKGLFFVRVOGYDNEYFMRAP	420
Qy	64	-----V-----S-I-----N-----ED-----GTLF-----	73
Db	421	AIGSVIUGPRAFWPIHQBFVGRDLNLSCTVESASAYTIYWKTEGDIIGGLFVHNTD	480
Qy	74	-----D-----G-----R-----P-I-----E-----	79
Db	481	TSVMTPELSLKDAGEYECRVISNNGNYSVKTRVETRESPEIFGVRNVSVPLGEAAFLH	540
Qy	80	-----S-----L-----SL-I-----DA-----	86
Db	541	CSTRSAGEVEIRWTRYGATVFNGETNPTNGLKIHVTRADAGVYECMARNAGMST	600
Qy	87	-----VM-----P-DV-----V-----	92
Db	601	RKMRLDIMEPPSVKVTPODYVFNMRGVNLSCEAMGDPKPEVHWYFKGRHLLNDYKYQVG	660
Qy	93	Q-----T-----R-----Q-----QAYRD-----KLA-----Q-----	105
Db	661	QDSKFLYIRDATHHDEGTECRAMSQAQA-RDITDLMLATPPKVEIIQNMWVGRDRV	719
Qy	106	-----S-----L-----Q-H-----A-----AAA-----	113
Db	720	SFECKTIRGKPKIRFKNGKOLIKPDYKIKNEGQLHMGAKDEDAAYSCVGENMAG	779
Qy	114	-----AAAA-----T-----N-----Q-----Q-GS-----A-----KN-----	126
Db	780	KDVQVANLSVGRVPTIIESPHTVRVNIERQVTLQCLAVGIPPPPEIEMQKGNVLLATLNP	839
Qy	127	-----GEN-----T-----A-N-----	132
Db	840	RYTQADG-NLLITDAQIEDQGOFTCIARTNYGQSQSOTLMTVGLVSPVLGHVPPBEOL	898
Qy	133	-----G-----G-----EE-----N-----G-----	137
Db	899	IEGQDLTLCVVVLGTPKPSIWIKDDKPVEEGPTIKIEGGSLRLRGPNPKDEKGYTC	958
Qy	138	-----A-----H-----TI-----A-----N-NH-----	145
Db	959	IAVSPAGNSTLHINVQLIKKPEFVYKPEGGIVFKPTISGMDEKHVAVVNSTHVDLDGEF	1018
Qy	146	-----T-----	146
Db	1019	AIPCWSGTPTIITWYLDGRPTPNRSDFTVTADNTLIVRKADKSYSGVYTCOATNSAG	1078
Qy	147	D-----M-----M-----E-----VD-----GD-----V-----	155
Db	1079	DNEQKTIIRIMNTFWISFGQSSFNWVDDLFTIPCDVYGPDPKPVITWLLDDKPPTEGVN	1138
Qy	156	E-----I-----PS-----N-----K-A-----VVL-----	165
Db	1139	EDGSLTIPNVNEAHRGTFTCHAQNAAGNDTRVTLTVHTTPTINAENQEKIALQNDDIVL	1198
Qy	166	-----R-----	167
Db	1199	ECPAKALPPPVRMLWYGEKIDSOLIPTHITREDGALVLQNVKLENTGVFCVQVSNLAGE	1258
Qy	168	-----HE-----SEY-----F-----I-C-A-----WNP-----V-----	180
Db	1259	SLSYTLTVHEKPKIISEVPGVVDVVKGTIEIPCRATGVPEVIRTNKNGIDLKWDEKKE	1318
Qy	181	S-D-L-L-V-S-----G-----SG-----D-----S-----TA-----	193
Db	1319	SYDNLGTLRIYEADKNDIGNYNCVVTNEAGTSQMTTHVDVQBPPIILPSTQTNNTAVGD	1378
Qy	194	RI-----W-----NL-----S-ENST-----S-----	204
Db	1379	RVELKCYVEASPPASVTWFRGIAIGTDTKGYVVEDSGTLVIOASVEDATITCKASNP	1438
Qy	205	-G-----P-----TQ-----L-----V-----L-----	211
Db	1439	AKAEANLOVTVIASPDIKDPDVVTQESIKESHPSLYCPVFSNPLPQISWYLNKPLID	1498
Qy	212	-----R-----H-----CI-R-----	216
Db	1499	DKTSWKTSDDKKLHVFKAKITDSGVYKCVARNAAGEGSKSPQVEVIVPLNLDSEKYYK	1558
Qy	217	-----EG-----G-----Q-----D-----V-P-----SN-----	225
Db	1559	VFAKGEFVTLGCPVSGFPVPQINWVDVTVPKPKYKATLSNDGLTLHFDVSUVKQE	1618
Qy	226	-----KVT-S-----L-----D-----W-----	233
Db	1619	GNVHCAQSKGNLIDIDVELSVLAVPIVGEDDNLVFLGKDISLSDLOTESDDKTFFW	1678
Qy	234	-N-SE-----G-----G-----TL-----L-----	240
Db	1679	SINGESDRPDNVQIPSDGHRLYITDAKPENNGKYMCRVNSAGKAERTLTLDVLEPPVF	1738
Qy	241	-----A-----TG-----SY-----D-----	246
Db	1739	VEPVFEANOKLIGNNPIILOQVTPGPKPTVIWKIDGNDVDKSWLFDSELSLLRIEKLIG	1798
Qy	247	-----G-----F-----A-----R-----I-----	251
Db	1799	KSAQISCTAENKAGTASRDFFIQNIAPTFKNEGDOETIFRESEITLDCPVSLGDFQIT	1858
Qy	252	W-----T-----K-----	255
Db	1859	WMKQGLPLTENDAIFTLDNTRLILNANRDHEDIYTCVANNTAGQVSKDFVQVQVLPKI	1918
Qy	256	-----GN-----L-----A-----S-----	260
Db	1919	KNAVTLNEEGEEIILTCDAENPTPTAKDFNQDLPKEAVFVNNHTVVNNVTKYH	1978
Qy	261	T-----L-----GQ-----H-----K-----	266
Db	1979	TGVYKCYATNKVQAVKTIINVHRTKPRFESGLTESELATVNLTRITLDCDVDDAIGVI	2038
Qy	267	-----G-P-----I-----F-----	270
Db	2039	SWTVNGKPLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVNEAGVATKTNLFVQVP	2098
Qy	271	-----AL-----K-----W-----	274
Db	2099	PTIVNEGGEYTVIENNSLVLPCVETGKPNPVVTWKDGRPVGLKSVQVLSGQGFQKIVH	2158
Qy	275	-----N-----K-----K-----	278
Db	2159	AEIAHGSYICMAKNDVGTAEISFDVDIITRPMIOKGINVITAIKGGALPFKCPIDDDK	2218
Qy	279	NF-----IL-----S-----AG-----	285
Db	2219	NFKGQIILWNYOPIDEADARITLSNDRRLTILNVNTEDEGOYSCRKNDAGENSFD	2278
Qy	286	-----V-----D-----KT-----TI-----I-W-----D-----A-H-----	296
Db	2279	FKATVLPPTIIMLDKDKKTAVEHSTVTLSCPATGKPEPDITWFKDGEAIIHENIADII	2338
Qy	297	-----T-----G-----EA-----K-----Q-----	302
Db	2339	PNGELANGNOLKITRIKEGDAGKYTCESADNSAGSVEQVNVNVTIPKIEKIDGPSYESQ	2398
Qy	303	Q-----	303
Db	2399	QNERVVISCPYARPPAKITWLKAGKPLQSDKFVKTSAHQKLVLPKLRDSSKYTCIA	2458

QY 304 -----F-----P-F-----H-----S----- 308  
Db 2459 TNEAGTKDRDFKYSMLVAPSFDEPNIVRRITVNSGNPSTLHCPAKGSPSTITWLKDGNA 2518  
QY 309 -----A-----P----- 310  
Db 2519 IEPNDRYVPFDAGRLQISKTEGSDQGRYTCIATNSVSGDDLENTLEVIIPPIDGERRE 2578  
QY 311 --A--L-D--VD--WO--S-N----- 319  
Db 2579 AVAVIEGFSSELCFDSNSTGVDEWQKDLGTINQDILRGDSFTQIPSSGKMSFLSARKS 2638  
QY 320 -----N-----T-----FA-----S-----C-S-T-- 327  
Db 2639 DSGRYTCIVRNPAGEARKLPDFAVNDPPSISDELSSANIQTIVPYYPVEINCVVSGSPHP 2698  
QY 328 -----DM-----C-----I- 331  
Db 2699 KVTYLFDDKPLEPDSAAAYELTNNGETLKIVRSQVEHAGTYTCEAONNVGKARKDFLRVT 2758  
QY 332 --H--V--C--KL--G----- 337  
Db 2759 APPHFKEREVARVGDTHMLTCNAESSVPLSSVYVWHAHDESVOQGVITSKYAANEKTL 2818  
QY 338 --Q-----Q-----DR--PI--K----- 343  
Db 2819 NVTNIQDDBEGFYCTAVNEAGITKKFKPLIVETPYFLDQOKLYPIILGKRLTDCSAT 2878  
QY 344 --T--F----- 345  
Db 2879 GTPPTILFMKQKRLNESDEVDIIGSTLVIDNPQKEVEGRYTCIAENKAGREKDMVVE 2938  
QY 346 -----Q-----G-H----- 348  
Db 2939 VLLPPKLSKEWINVEVQAGDPLTECPIDETSGVHITWSRQFGKQDOLMRAQSSDKSK 2998  
QY 349 --T--N--E--VN-----A----- 354  
Db 2999 LYIMQATPEDADSYCIAVNDAGAEAVQVTVNTPPKIFGDSFSTTEIVADTLEIPCR 3058  
QY 355 -----IKW-----D 358  
Db 3059 TEGIPPEISWFLDGKPILEMPGVYKQGLSLRIDNIKPNQEGRYTCVAENKAGRAEQD 3118  
QY 359 -----P-----T-----GN----- 362  
Db 3119 TYVEISEPPRVNMASEVMRVVEGRQTTIRCEVFNPEPVVNNLKGEPYTSDDLQFSTKL 3178  
QY 363 --L--LA--SC--S--D--D-----D-----M-- 371  
Db 3179 SYLHLRETTLDGTYTCIATNKAGESOTTDDVEVLVPPRIEDEERVLOKSGNTYVHVC 3238  
QY 372 -----L-----T-----L----- 373  
Db 3239 QVTGRPVYVTKRNGKEIQFNPVLHNRATRADEGKYSCIASNEAGTAVADFLIDVFT 3298  
QY 374 -----KI-----W-----S-----M--K 379  
Db 3299 KPTFETHETTFNIVEGESAKIECKIDHPKPTISMLKGRPFNMDNIILSPRGDTLMILK 3358  
QY 380 -----QD--N-----C----- 383  
Db 3359 AORFDGGLYTCVATNSYSGSEQDFKNVYTKPYVIDETIDQTPKAVAGGEIILKCPVLGNP 3418  
QY 384 -----V--H--DL-----Q-----A-----H-NKE-I 394  
Db 3419 TPTVTWKRGDDAVPNDSTRHTIVNNYDLKINSVTTEDAGQVSCIAVNEAGNLITHYAAEVI 3478  
QY 395 -----Y-----TI-K-----W-----SP----- 401  
Db 3479 GKPTFVRKGNLVEIENDTITMDCGVTGRPLPSISWFRGDKPVYLYDRYSISPDGSHIT 3538

QY 402 -----T----- 402  
Db 3539 INKAKLSDGKYICRASNEAGTSDIDLILKILVPPKIDKSNIIGNPLAIVARTIYLECPI 3598  
QY 403 -G--P-----G--T-----NN----- 408  
Db 3599 SGIPODPVITWTKNGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGKASHD 3658  
QY 409 -----P----- 409  
Db 3659 FSLDVLSPPEPDHGTQPTIKREGDTITLTCPKLAEDIADQVMDVSWTKDSRALDGLT 3718  
QY 410 -N-----A--N--L--ML--A-SAS--F-----D-- 421  
Db 3719 DNVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPVVIDISRNVDVQPQ 3778  
QY 422 -----S-----T--VR-----L-----W-- 427  
Db 3779 VAVNQPTIMRCAVTHGHPFSIKWLKNGKEVTDDENIRIVEQGVQLILRTDSHPAGKWSC 3838  
QY 428 -----DV-----D--R--G--I--C----- 434  
Db 3839 VAENDAGVKELEWLVDFTPPVVSVKSDNPICALGETITILFCVASGNPYPOLKWKAGGSL 3898  
QY 435 -----I-H-----T----- 437  
Db 3899 IFDSPDGAISLKGARLDIPHLKKTVDVGTQCALNAAGTSEASVSDVLVPPPEINRDCI 3958  
QY 438 --JT--K-----H-----Q----- 442  
Db 3959 DMSPLPAQOQSLTLOCLAQKPVQPMRWTLNGTALTHTSPGITVASDSTFIQINNVSLS 4018  
QY 443 -----E-----PV-----Y-----S-- 447  
Db 4019 KGVYTCYENAVAGSNLMTYNDVVOQAPVISNGTKQVIEGELAVIECLVEGYPAQVSWL 4078  
QY 448 -----V-----A----- 449  
Db 4079 RNGNRVETGVQVRYVTDGRMLTIIEARSLDSGIVLCSATNEAGSAQAQVTLVLPVKI 4138  
QY 450 -----FS-----P-----DG----- 454  
Db 4139 ITSTPGVLPSSGSKFSLPCA VRGYPDPPIISWTLNGNDIKOGENGHTIGADGTLHIEAE 4198  
QY 455 -R--Y-----L-----A--SG----- 460  
Db 4199 ERHLIYECTAKNDAGADTLEFPVQTVIVAPKISTSGNRYINGSEGTETVIKCEIESSESEF 4258  
QY 461 -----S--F--D--K-----C----- 465  
Db 4259 SMSKNGVPLLPNNLIFSEDYKLIKILSTRSDQGSYSCTAANKAGNATQKTNLNVGVAP 4318  
QY 466 --VH--I-W-----NT----- 471  
Db 4319 KIMERPTQVHKGDQVTLWCESAGVPQPAITYKDNELLTNTGTVDDETATTKKSVIFSS 4378  
QY 472 -----TG--Q----- 474  
Db 4379 ISPSQAGVYTCANWVASTESDIDLIVMIPPEVVPERNVNTNPRQTVFLSCNATGIPE 4438  
QY 475 -----A-----L--V-----H----- 478  
Db 4439 PVISMWRDSNIAIQNNEKYQILGTLAIRNVLPPDDGFYHCIAKSDAGQKATKRLIVNK 4498  
QY 479 -----S--YR--G-- 482  
Db 4499 PSDRPAPIWVECEKGPKKTEYIMIDRGDTPDDNPOLLPWKQVEDSSLSNGSIAYRCMPGP 4558  
QY 483 -----T-----G-----G--I-----FE-- 488  
Db 4559 RSSRTVLLHAAPQFIVKPKNTAAI GAIVELRCSAAGPPHPTITWAKDGKLEDSKFEIA 4618  
QY 489 -----V--C-- 490

Db 4619 YSHLKVTLNSTSDSGEYVMAQNSVGSSTYSAFINVDNNILPTPKSSNQKNVAVITCYE 4678  
 QY 491 -----W-----N-A----- 493  
 Db 4679 RNOAYSRLTWEYNGVPMPLAGIHPMNGSLVILDTSSLKEGDLELYTCKVRNRRHS 4738  
 QY 494 -----A-G-----DKV-----G-----AS-----A--- 502  
 Db 4739 IPHLTSAFEGVPEVKTIDRKEVNGDSVLDCVTSPLTHVVTKNQDKMLDDDAIYV 4798  
 QY 503 -----S-----D-GSV-CV-----LD-----L-----R-----K 514  
 Db 4799 LPNNSLVLLNVEKYDEG-VYKVASNSIGKAFDDTQLNVYVGGSSREAYK 4847

RESULT 2  
 T43290  
 hemicentin precursor - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 09-Jul-2004  
 C:Accession: T43290; T20993; T24734  
 R:Vogel, B.E.; Hedgecock, E.M.  
 A:Description: Hemicentin is required for hemidesmosome mediated cell adhesion and germ-  
 submitted to the EMBL Data Library, June 1998  
 A:Reference number: 222396  
 A:Accession: T43290  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-5198 <W0>  
 A:Cross-references: UNIPROT:O76518; EMBL:AF074901; PIDN:AAC26792.1  
 R:Sulston, J.  
 submitted to the EMBL Data Library, December 1994  
 A:Reference number: Z19355  
 A:Accession: T20993  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-5198 <W1>  
 A:Cross-references: EMBL:Z47068; PIDN:CAA87336.1; GSPDB:GN00028; CBSP:F15G9.4b  
 A:Experimental source: clone F15G9  
 R:Kershaw, J.  
 submitted to the EMBL Data Library, December 1994  
 A:Reference number: Z19929  
 A:Accession: T24734  
 A:Status: preliminary; translated from GB/EMBL/DDBJ  
 A:Molecule type: DNA  
 A:Residues: 1-5198 <W2>  
 A:Cross-references: EMBL:Z47070; PIDN:CAA87345.1; GSPDB:GN00028; CBSP:F15G9.4b  
 A:Experimental source: clone T09B9  
 C:Genetics:  
 A:Gene: him-4; F15G9.4b  
 A:Map position: X  
 A:Introns: 85/1; 120/1; 334/3; 370/1; 477/2; 606/3; 664/1; 935/3; 977/1; 1051/3; 1184/3;  
 2512/2; 2593/3; 2699/3; 2759/1; 2852/1; 2889/3; 2913/3; 2941/1; 2967/3; 2991/3; 3033/1;  
 1; 4225/1; 4361/1; 4408/1; 4456/1; 4498/1; 4647/3; 4838/1; 4902/1; 4964/1; 5034/1; 5100/1

Query Match  
 Best Local Similarity 71.2%; Score 2610.8; DB 2; Length 5198;  
 Matches 447; Conservative 52; Mismatches 12; Indels 4362; Gaps 387;  
 QY 1 M-----SI-----SS-----D-----E----- 7  
 Db 1 MGRSPSWLYGVLLGALLATTCSSVNDKNDKNDPTCKSSIAFVFDITGSMFDDLVQVREGAAK 60  
 QY 8 -----V-----N-----F-----L-----VY----- 13  
 Db 61 IFKTVMAQREKLIYNIWPFHPDLPYLGEIINTTDTSTYFMRQLSKVYVHGGDCPEKTLTG 120  
 QY 14 -----R-----Y-----L-----QE-----SG-----F-----SHSAF-T 27  
 Db 121 ILKALQISLSPSSYIYFTDARSKDYLEDEVLNTIQEKSSVVFVMTGCGNETHPGFRT 180  
 QY 28 -----FG-I-----E-----S-----H-ISQ-----S-NI----- 39

Db 181 YEKIAAASFGQVPHLEKSDVSTVLEVRHAVKQKVHLMAYEARERGGTVSRNIPVDKHL 240  
 QY 40 -----N-----GA-----LV-----P-P----- 46  
 Db 241 ELTISLSDGKDDSDNLDIVLRDPEGRVTKRLYSKEGGTIDLKNVKLIRLKDPSPGVTV 300  
 QY 47 -----A-----A-----A-----L-I-----S-I----- 52  
 Db 301 NNTNSRLKHTIRVFGHGAVDFKYGFAISRPLDRIELARPRPVNLQDVTLLINMTGLIPPGTV 360  
 QY 53 --I-----Q-----KGL-----Q-Y-----VE-----AE----- 63  
 Db 361 GEIDLVDYHGHSLYKAVASPHRTNPNMYFAGPPVPKGLFFVRVQGYDENDYEFMRIA 420  
 QY 64 -----V-----S-I-----N-----ED-----GTLF----- 73  
 Db 421 AIGSVIVGGPRAPFMSPIHQEFVGRDLNLSCTVESASAYTIYWKGTEDIIGPLFYHNTD 480  
 QY 74 -----D-----G-----R-----P-I-----E----- 79  
 Db 481 TSVWTTPELSLKDAGEYECRVISNNGNYSVKTRVETRESPPPELFGVRNVSVPLEAFLH 540  
 QY 80 -----S-----L-----SL-I-----DA----- 86  
 Db 541 CSTRSAGEVEIRWTRYGATVFNCPNTPNRTNGTLKIHVTRADAGVYECMARNAGMST 600  
 QY 87 -----VM-----P-DV-----KLA-----Q-----V----- 92  
 Db 601 RKMRLDIMEPPSVKVTPODYFNMREGVNLSCAMGDPKPEVHWYFKGRHLLNDYKYQVG 660  
 QY 93 Q-----T-----R-----Q-----QAYRD-----KLA-----Q----- 105  
 Db 661 QDSKFLYIRDATHDEGTCECRAMSQAGQA-RDITDMLATPPKVEIIQNMVMVGRDRV 719  
 QY 106 -----AAA-----Q-H-----A-----AAA-----AA----- 113  
 Db 720 SPECKTIRKPHKIRWFKNGDLIKPDDYIKINEGQLHMGAKDEADAGAYSCVGENMAG 779  
 QY 114 -----AAAA-----T-----N-----Q-----Q-GS-----A-KN----- 126  
 Db 780 KDQVQANLSVGRVPTTIESPHTVRVNIERQVTLQCLAVGIPPEIEMQKGNVLLATLNNP 839  
 QY 127 -----GEN-----T-----A-N----- 132  
 Db 840 RYTQLADG-NLLITDAQIEDQOFTCIARTYQQSQSTLMTVGLVSPVLGHVHPPEEQ 898  
 QY 133 -----G-----EE-----N-----G----- 137  
 Db 899 IEGQDLTLCVVVLTGTPKPSIVWIKDKPVEEGPTIKIEGGSLRLRGNGPKDEKGYTC 958  
 QY 138 -----A-----H-----TI-----A-N-NH----- 145  
 Db 959 IAVSPAGNSTLAINVOLIKKPEFVYKPEGGIVFKPTISGMDEKHAVAVNSTHDLVDEGF 1018  
 QY 146 -----T----- 146  
 Db 1019 AIPCVCVSGTPPPIITWYLDGRPIITNSRDFVTADNTLIVRKADKSYSGVYTQATNSAG 1078  
 QY 147 D-----M-----E-----VD-----GD-----V----- 155  
 Db 1079 DNEQKTTIRIMNTPMISPGQSSFNMMVVDLFTIPCDYGDPKPVTIWLDDDKPFTGEGVN 1138  
 QY 156 E-----I-----PS-----N-K-A-----VVL----- 165  
 Db 1139 EDGSLTIPNNEAHRGTFTCHAQNAAGNDTRVTLTPTTINAENQEKIALQNDIVL 1198  
 QY 166 -----R-----G----- 167  
 Db 1199 ECPAKALPPPRVLTWYEGEKIDSQIPHTIREDGALVQLONVKLENTGVFCVQVSNLAGED 1258  
 QY 168 -----HE-----SEV-----F-----I-C-A-----WNP-----V----- 180

Db 1259 SLSYTLTVHEKPKIIEBPGVWDVKGFTIEIPCRATGVPEVIRTNKNGIDLKMDEKFF 1318  
QY 181 S-D-L-L-V-S-G-SG-D-S-TA- 193  
Db 1319 SVNGLTLRIEADKNDIGNYCVVNEAGTSOMTHVDVQBPPIILPSTQTNNTAVVGD 1378  
QY 194 RI-W-NL-S-ENST-S- 204  
Db 1379 RVELKCVASPPASVTFRRGIAIGTDTKGVVVEDGLTVIQSASVEDATIYCKASNP 1438  
QY 205 -G-P-TQ-L-V-L- 211  
Db 1439 AGRAEANLQVTVIASPDIKDPDVVTQESIKESHPSLYCFVFNPLFOISWYLNDRKPLID 1498  
QY 212 -R-H-CI-R- 216  
Db 1499 DKTSWTSDDKRLHVFKAKITDSGVYKCVARNAAGEGSKSFQEVIVPLNLDESQYKKK 1558  
QY 217 -EG-G-Q-D-V-P-SN- 225  
Db 1559 VFAKEGEVTLGCPVSGFPVQINWVDGTVVEFGKYGATLSNDGLTLHFDVSQOE 1618  
QY 226 -KVT-S-L-D-W- 233  
Db 1619 GNYHCVAQSGKNILIDVLSVLAVPIVGDDNLEVLFGKDISLSCDLQTESDDKTTFW 1678  
QY 234 -N-SE-TL-L- 240  
Db 1679 SINGESDRPDVQIIPSDGHRLYITDAKPENNGKYMCRVNSAGKAERLTLVDLEPPVF 1738  
QY 241 -A-TG-SY-D- 246  
Db 1739 VEPVFEANQKLGNNPIILQCVQTNPKPTVIWKIDGNDVDKSWLFDSELSLLRIEKLGT 1798  
QY 247 -G-F-A-R- 251  
Db 1799 KSAQISCTAENKAGTASRDFQIONAAPTFRNEGDOETIFRESEITLDCPVSLGDFQIT 1858  
QY 252 W-KD- 255  
Db 1859 WMKQGLPLTENDAIPLDNLTRLILNANRDHDIYTCVANNTAGQVSKDFDQVQVLPKI 1918  
QY 256 -GN-L-A-S- 260  
Db 1919 KNAVVTLEINEGEEIILTCDAENPTPTAKWDNFQGDLPKEAVFVNNHTVVNNVTKYH 1978  
QY 261 T-L-Q-Q-H-K- 266  
Db 1979 TGVIKCYATNKVQAVKTINNVHRTKPRFESGLTESELTVNLFRSITLBCDVEDDAIGVGI 2038  
QY 267 -G-P-I-F- 270  
Db 2039 SWTVNGKPLAETDGVQTLAGGRFLHIVSAKTDHGSYACTVNEAGVATKTNLFVQVP 2098  
QY 271 -AL-K-W- 274  
Db 2099 PTIVNEGGEVTVIENSLVLPCEVTGKPNPVVTTKDRPVGLKSVQVLSQGFQKIVH 2158  
QY 275 -N-K-KG- 278  
Db 2159 AEIAHGSYICMAKNVDGTAIESFDVDIITRPMIOKGIKNIVTAIKGGALPFPKCPIDDDK 2218  
QY 279 NF-IL-S-AG- 285  
Db 2219 NFKGQIILNRYOPIDLEADARITLSNDRRLTLNVNTEDEGGQYSCRVDNAGNSFD 2278  
QY 286 -V-D-KT-TI-I-W-D-A-H- 296  
Db 2279 FKATLVPPPIIMLDKDKNKAVEHSTVLSCTATGKPEPDITWFKDGEAIIENIADII 2338  
QY 297 -T-G-BA-K- 302  
Db 2339 PNGELNGNQLKIPRIKEGDAKGYTCESADNSAGSVEQDVNNVNVTIPKIEKGIPSDYESQ 2398

QY 303 Q- 304  
Db 2399 QNERVVISCPVYARPPAKITWLKAGKPLQSDKFKVTSANGCKLYLFLKRETDSSKYTCIA 2458  
QY 305 -P-F-H-S- 308  
Db 2459 TNEAGTKRDFKVMVAPSFDEPNIVRRITVNSGNPSTLHCPAKGSPSPITWLKDGNA 2518  
QY 309 -A-P- 310  
Db 2519 IEPNDRYVFDAGRQLQISKTBSGQGYTCIATNSVGDDLENTLEVIIPPIDGERRE 2578  
QY 311 -A-L-D-VD-WO-S-N- 319  
Db 2579 AVAVIEGFSSELCFDSNSTGVDVWQKDGTLINQDTRLGDSFIQIPSSGKMSFLSARKS 2638  
QY 320 -N-T-FA-S- 327  
Db 2639 DSGRYTCIVRNPAAGEARKLFDFAVNDPPSISDELSSANIQTIVPYYPVEINCVSQSGPH 2698  
QY 328 -DM- 331  
Db 2699 KYVWLFDDKPLBPSAAVELTNGETLKIIVRSQVEHAGTYTCEAQNNGKARKDFLVRVT 2758  
QY 332 -H-V-C-KL-G- 337  
Db 2759 APHPKEKEREVVARVGDTHMLTCAESSVPLSSVYVWHAHDESVQNGVITSKYAANEKTL 2818  
QY 338 -Q-DR-P-K- 343  
Db 2819 NVTNIQLDDGFGYCTAVNEAGITTKFKFLIVETPYFLDQKLYPIILGKRULDCSAT 2878  
QY 344 -T-F- 345  
Db 2879 GTPPPILFMKDGKRLNESDEVDIIGSTLVIDNPQKEGRYTCIAENKAGSEKMMVE 2938  
QY 346 -Q-G-H- 348  
Db 2939 VLLPPKLSKEWINVEVQAGDPLTECPIEDTSGVHITWSRQFGKQGLDMRAQSSDQSK 2998  
QY 349 -T-N-E-VN-A- 354  
Db 2999 LYIMQATPEDADSYSCIAVDAGAEAVFQVTVNTPPKIFGDSFSTTEIVADTLEIPCR 3058  
QY 355 -IKW- 358  
Db 3059 TEGIPPEISWFLDGKPILEMPGVTYKQGLSLRIDNIKNQEGRYTCVAENKAGRAEQD 3118  
QY 359 -P-T-GN- 362  
Db 3119 TYVEISEPVRVMASEVMRVVEGRQTTIRCEVFGNPEPVVNWLDGEPYTDLLQFSTKL 3178  
QY 363 -L-LA-SC-D- 371  
Db 3179 SYLHLRETTLAGDGTCTIATKNAGESQTTDVEVLVPPRIEDEERVLOGKEGNTVMVHC 3238  
QY 372 -T-L- 373  
Db 3239 QVTGRPVVYVWKRNGKEIEQPNVLHNRATRADGKYSCTASNEAGTAVADFLIDVFT 3298  
QY 374 -KI-W-S-M-K- 379  
Db 3299 KPTFETHETFNIVEGESAKIECKIDGHPKPIISMLKGRPFNMWNIILSPRGDTLMILK 3358  
QY 380 -QD-N- 383  
Db 3359 AQRFDDGLYTCVATNSYGDSEQDFKVVVTKPVIDBTIDQTPKAVAGGBIILKCPVLGNP 3418  
QY 384 -V-H-DL-Q-A-H-NKE-I- 394  
Db 3419 TPTVTWKRGDDAVPNDRSRHTIVNNDYDLKINSVTTEDAGQYSCIAVNEAGNLTHYAAEVI 3478

QY 395 -----Y-TI-K-----W-----SP-----401  
Db 3479 GKPTVRKGNLYEVIENDTTIMCGVTSRPLPSISWFRGDKPVLYLDRYSISPDGSHIT 3538  
QY 402 -----G-----P-----T-----402  
Db 3539 INKAKLSDGGKVICRASNEAGTSDIDLILKILVPPKIDKSNIGNPLAIVARTIYLCPI 3598  
QY 403 -G--P-----G-----T-----NN-----408  
Db 3599 SGIPQPDVITWTKGMDINMTDSRVILAQNNETFGIENVQVTDQGRYTCTATNRGKASHD 3658  
QY 409 -----S-----P-----409  
Db 3659 PSLDVLSPPEPIHGTOPTIKREGDTTLTCTPIKLAEDIAQVMDVSWTKDSRALDGLT 3718  
QY 410 -N-----A--N--L-----ML--A--SAS--F-----D-----421  
Db 3719 DNVDISDDGRKLTISQASLENAGLYTCIALNRAGEASLEFKVEILSPVIDISRNDVQPO 3778  
QY 422 -----S-----T-----VR-----L-----W-----427  
Db 3779 VAVNQPTMCAVTHGPFPPSKLWKGKVEVTDENIRIVEQGVQLQILRTDSHAGKWC 3838  
QY 428 -----DV-----D--R--G--I--C-----434  
Db 3839 VAENDAGKLEWLDVFTPPVSVKSDNPIKALGETITILFCNASGNYPQLKWKAGGSL 3898  
QY 435 -----I--H-----T-----437  
Db 3899 IFDSPDGARISLKGARLDIPLHKKTDVGDYTCQALNAAGTSEASVSDVLVPPPEINRDI 3958  
QY 438 -----LT-----K-----H-----Q-----442  
Db 3959 DMSRPLPAQOSLTLLQCLAQKQPPQMRWTLNGTALTHTPGITVASDTPFIQINNVSLS 4018  
QY 443 -----E-----PV-----Y-----S--447  
Db 4019 KGVVTCVAENVAGSDNLMNVVVQAPFVINSNGGTVQVIEGELAVIECLVEGYAPQVSWL 4078  
QY 448 -----V-----A-----449  
Db 4079 RGNRNVETGVGVYVTDGRMLTIIARSLSLDSGLYLCSATNAGSAQAQVTLVLSPKI 4138  
QY 450 -----FS-----P-----DG-----454  
Db 4139 ITSTPGVLTSPSSGKSLPCAVRGYDPPIISWTINLNDIKOGENHTIGADGTLHIEKAE 4198  
QY 455 -R--Y-----L-----A-----SG-----460  
Db 4199 ERHLIVECTAKNDAGADTLEFPVQTVIVAPKISTSGNRYINGSGTETVIKCEIESSESF 4258  
QY 461 -----S--F--D--K-----C-----465  
Db 4259 SWSKNGVPLPNSNLIIFSEDKILKILSTRLSDQGEYSCTAANKAGNATQKTNLVGVAP 4318  
QY 466 -----VH-----I--W-----NT-----471  
Db 4319 KIMERPTQVHKGDQVTLWCEASGVPOPAITWYKDNELLTNTGVDETATTKKSVIFSS 4378  
QY 472 -----Q-----TG-----474  
Db 4379 IPSQAGVYTCKAENWVASTEEDIDLIVIMPEWPMNMVSTNPTQTVFLSCNATGPE 4438  
QY 475 -----A-----L--V-----H-----478  
Db 4439 PVISWRDSNIALONNEKYQILGTTLAIRNLVDDGDFHCTAKSDAGOKIATKRLIVNK 4498  
QY 479 -----S--YR--G--482  
Db 4499 PSRRPAPIWECDEKGPKTEYMWIDRGTPDDNPQLLPWKVDSDSLNGSIAYRCMPGP 4558  
QY 483 -----T-----G--I-----FE--488

Db 4559 RSSRTVLLHAAPOFIIVKPKNTTAAIGAIVELRCSAAGPPHTTITWAKDGKLIEDSKFEIA 4618  
QY 489 -----V--C--490  
Db 4619 YSHLKVTLNSTSDGEYTCMAQNSVGSSTVSFAFVNDNNILPTPKPSSNQKNVAVITCYE 4678  
QY 491 -----W-----N--A-----493  
Db 4679 RNOAYSRLTWEYNGVMPKPNLAGIHFMNNGSLVILDTSSLKEGDLELYTCKVNRNRHS 4738  
QY 494 -----A--G-----DKV-----G-----AS-----A--502  
Db 4739 IPHLTSAFEGVPEVKTIQVNVNNGSVVLDCEVTSDELTTTHVVVTKNDQKMLDDDAIYV 4798  
QY 503 -----S-----D--GSV--CV-----L-----D--L-----512  
Db 4799 LPNNSLVLLNVEKYDEG--VYKCVASNSIGKAPDDTQLNVEYEGDFLPTGTGEGSGINIDDS 4857  
QY 513 -----R--K 514  
Db 4858 SNAGSSRREAYK 4870

## RESULT 3

I38346

elastic titin - human (fragment)

C;Species: Homo sapiens (man)

C;Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 09-Jul-2004

C;Accession: I38346

R;Label: S.; Kolmerer, B.

Science 270, 293-295, 1995

A;Title: Titins: Giant proteins in charge of muscle ultrastructure and elasticity.

A;Reference number: A57430; MUID:96026330; PMID:7569978

A;Accession: I38346

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-7962 &lt;RES&gt;

A;Cross-references: UNIPROT:Q10465; EMBL:X90569; NID:g1017426; PID:CAA62189.1; PID:g101

C;Genetics:

A;Gene: GDB:TTN

A;Cross-references: GDB:127867; OMIM:188840

A;Map position: 2q31-2q31

## Query Match

Best Local Similarity 71.0%; Score 2603.7; DB 2; Length 7962;

Matches 448; Conservative 52; Mismatches 11; Indels 4663; Gaps 377;

QY 1 MS-----IS-----SDE-----V-N-----9  
Db 57 MTQPEDTYQLEIAEAYPEDEGTYTFVANNVQGVSTANLSLEAPESILHERIEQIEME 116  
QY 10 -----F-----L-----VY-----R--Y--15  
Db 117 MKAAPVKKRKBLEVALGHAKFTCEIQSAPNRFQWFKAGREIVESDKSIRSKYIS 176  
QY 16 ---L--Q-----E-----S-----G--P--21  
Db 177 SLEILRTQVDCGEYTCASNEYSVSCATLTVTVEAYPTFLSRPKSLTTFVGVKAAPFI 236  
QY 22 -----S-----H-----S-----24  
Db 237 CTVTGTPVETIWKDGAALSPNWRISDAENKHLISLNTIQDRGVYSCAKNFKGA 296  
QY 25 ---A---P-----T-----27  
Db 297 DICQAEIIIDKPHFIKELEPVQSAINKKVHLECOVDEDRKVTVTWSKDGKLPKPKDYK 356  
QY 28 --F-----G-----I-----30  
Db 357 ICFEDKIATLEIPLAKLXDSGYTCTASNEAGSSCSATVTVREPPSVFKKVDPSYMLP 416  
QY 31 --ES--H-----ISQ-----SN-----G-----41



Db 417 GESARLHCKLKGSPVI-QVTFKNNKELGESNTVRMFVNSEAILDITDVKVEDSGSYSC 475  
QY 42 -A- -LV- -P- -DA- -AL- -IS- -51  
Db 476 EAVNDVGSDCSTEIVIKPPSPFIKLEPADIVRGTNALLQCEVSGTGPFBSWFKDKKQ 535  
QY 52 -I- -52  
Db 536 IRSSKKYRLFSQSLVCLBIFSPNSADVGECVAVNEVGKCGCMATHLLKEPPTFKVKV 595  
QY 53 -I- -Q- -KG- -L- -Q- -58  
Db 596 DDLIALGGOTVTLQAAVRGSEPISTVMRGQEVIREDKIKMSFNGVAVLIIPDVQISF 655  
QY 59 -Y- -VE-AE- -63  
Db 656 GGYTCLAENAGSQTSGELIVKEPAKIERAELIQTAGDPATLEYTVAGTPELKPWK 715  
QY 64 -V-S- -I- -N- -E-D- -G- -T-L- -72  
Db 716 YKDRPLVASKKYRISFKNNVAQLKFYSALHDSGOYTPEISNEVGSSCETFTFVLDRD 775  
QY 73 -P- -D- -G-R- -P- -I- -E-SL 81  
Db 776 IAPFFTKPLRNVDVVNGTCRLDCKIAGSLPMRVSWFKDKKEITAASDRYIAFVEGTASL 835  
QY 82 SLI- -DA- -V-MP- -DV- -V- -Q 93  
Db 836 EIIIVDMNAGNPTCRATNSVGSKDSGALIVQEPSPFVTKPGSKDVLPGSAVCLKSTFQ 895  
QY 94 -T-R- -95  
Db 896 GSTPLTIRWFGKNKELVSGSGYITKEALESLELVKTSDSGYTCVSNVAGGVCS 955  
QY 96 -Q- -Q-A- -Y- -99  
Db 956 ANLFVKEPATFVEKLEPSQLKKGDATOLACKVTGTPPIKITWAFANDREIKESKHMFSF 1015  
QY 100 -R- -D- -KL- -AQ- -105  
Db 1016 VESTAVLRITDVGIEDSGBYMCEAQNAGSDHCSSIVIVKESPYTFKEKPIEVLEKVDV 1075  
QY 106 -AA- -Q-H- -AA-A- -110  
Db 1076 MLLAEVAGTPPEITWFKDNTILRSGRKYKTFIQDLHLSLQILKFVAADAGEYQCRVYNE 1135  
QY 111 -AA- -AAA-AAT- -118  
Db 1136 VGSSICARVTLREPPSPFIKKIESTSSLRGGTAAFQATLKGSLPITVTWLKDSDEITDD 1195  
QY 119 -N- -Q- -120  
Db 1196 NTRMTFENNVAASYLSGIEVKGDKYVCOAKNDAGIQRCSALLSVKEPATITTEAVSIDV 1255  
QY 121 -Q- -S- -AK- -125  
Db 1256 TOGDPAVLQVKSQGTKEITAKWFKQGBELTGLSKYKISVTDVTSILKIISTEKDGSGET 1315  
QY 126 -N- -G- -127  
Db 1316 FEVQNDVGRSSCKARINVLDLIIPSPFTKKLKMDSIKGSFIDLECIVAGSHPIQWPK 1375  
QY 128 -E- -NTA- -N-G- -E- -134  
Db 1376 DQOEISASKYKPSFHDNTAFLEI1SOLEGTDSGTYTCSATNKAHNCQSGHLTVKEPPYF 1435  
QY 135 -E- -N- -G-A- -H- -139  
Db 1436 VEKPSQDVPNTRVQLKALVGTAPMTIKWPKDKNELHSGAARSVWKDDTSTSLFPA 1495  
QY 140 -T-I- -A- -N- -143

Db 1496 KATDSGTIICQLSNDVGTATSKATLFFVKBPPOFIKPPSPVLVRNQSTTFECQITGTPK 1555  
QY 144 -N- -H- -T- -D- -147  
Db 1556 IRVSWYLDGNEITATQKHGISFIDGLATFQISGARVENSGTYVCEARNDA GTASCSELK 1615  
QY 148 -MM- -EV- -D- -G- -D- -V- 155  
Db 1616 VKEPPTFIRELPVVEVKYSDVELECEVTGTPPFVTLKNNREIRSSKKYTLTDVRSVF 1675  
QY 156 -E- -I- -PS- -N- -K- -161  
Db 1676 NLHITKCDPSDTGEYQCIVNEGGSCSTRVALKEPPSFIKKIENTTTVLKSSATFQST 1735  
QY 162 -A- -V- -L- -R- -CH- -ES- -170  
Db 1736 VAGSPISITWLKDDOILDEDDNNVISFVDSVATLQIRSDNGHSGRYTCQAKNESGVER 1795  
QY 171 -170  
Db 1796 CVAFLLVQEPAQIVEKAKSVDTVTEKDPMTLECVVAGTPELKVWLKDGKQIVPSRYFSMS 1855  
QY 171 -E- -V- -P- -173  
Db 1856 FENNVASPRIOSVMKQDSQYTFKVENDFGSSCDAYLRVLDONIPPSFTKKLTMKDKVL 1915  
QY 174 -I- -C- -A-W- -177  
Db 1916 GSIHWECKVSGSLPISAQWFKDGKEISTSAKRYLVCHERSVSVLENNLEBTDANYTCK 1975  
QY 178 -N- -P- -V- -S- -G-DS- -T- -DL-L 184  
Db 1976 VSNVAGDADACGILTVKEPPSFLVKPGRQQAIPDSTVEFKAILKGTPTPKIKWFKDDVEL 2035  
QY 185 VSG- -S- -G-DS- -T- -192  
Db 2036 VSGPKCFLEGSTFNLVSYVDASKTGOYTCHVTNDVGDSDCTMLLVTEPPKFKVCKLE 2095  
QY 193 -A- -R- -I- -W- -NLS- -199  
Db 2096 ASKIVAGDSSLECKIAGSPEIRVVWFNEHELPA SDKYRMTFIDSVAVIQNNLISTED 2155  
QY 200 -E- -N- -ST- -SG- -P- -T- 207  
Db 2156 SGDFICEAQNAGTSCTSKVIVKBPFPFPIVETLKNABVSLECELSGTPPEVVM 2215  
QY 208 -QL- -VL- -RHC- -214  
Db 2216 YKDKRQLRSSKKYKIAKNFHTSIHILNVDTSDIGYHCKAQNEVGSDTCVCTVKLKEPP 2275  
QY 215 -IRE- -17  
Db 2276 RFVSKNLSLTVVAGBPAELQASIEGAQPIFVQWLKEKEVIRESENIRITFVENVATLOF 2335  
QY 218 -GG- -219  
Db 2336 ACAEPANAGKYICQIKNDGGRNNMATLWLEPAVIVEKAGPMTVTGCTCLECKVAGT 2395  
QY 220 -OD- -221  
Db 2396 PELSVEWYKDGKLLTSSQKHKFSFYNNKISSRLSILSVEROAGTYTFVQVQNNVKSCTAV 2455  
QY 222 -V-PS- -N- -K- -226  
Db 2456 VDVSDRAVPPSTRRLKNTGGVGLGASCILECKVAGSSPISVAFWEHKTIVSGAKYQTFP 2515  
QY 227 -D-V-T- -SLD- -232  
Db 2516 SDNVCTQLNSLSDSDMGNYTCAANVAGSDSCRAVLTVQEBPPSFVKEPEPELVLPKQV 2575  
QY 233 -W- -N- -S-E- -236  
Db 2576 TFTSVIRGTPPKVNNFRGARELVKGDRCNIYFEDTVAELELFNIDISQSGEYTCVVSNN 2635

QY 237 -G---T-LL---A-----TG---S--Y--DGF----- 248  
Db 2636 AGQASCTRLRVKEPAAFLKRLSDHSVEPGKSIILESTYGTGLPISVTWKXGDNITSE 2695  
QY 249 -----A-R----- 250  
Db 2696 KCNIVTTEKTCILILNSTKRDAGQYSCIEINEAGRDVCGALVSTLPPYFVTEPLEA 2755  
QY 251 -----I--W---TK-----D-G--- 256  
Db 2756 AVGDSVSLQCVAGTPEITVSWYKGTCLRPTPEYRTYFTNNVATLVFNKVNINDSGEYT 2815  
QY 257 ---N-L--AS-----TL---G---Q----- 264  
Db 2816 CKAENSGTASSKTVPRIQERQLPPSFARQLKIEQVGLPVTUTCLRLNGSAPIQVCWYR 2875  
QY 265 -----H---K-----G----- 267  
Db 2876 DGVLLRDHENLQTSFVDNVATLKIQLDLSHGQYSCSASNPLGTASSSARLTAREPKS 2935  
QY 268 ---P---I---F---A--LK--W--NK----- 276  
Db 2936 PFDIKVSDIVIAGESADFECHVTGAQPMRITWSKDKNEIRPGNYTITCVGNTPHLRI 2995  
QY 277 ---KG----- 278  
Db 2996 LKVGKDSGGYTCQATNDVGDKMCQAQLSVKEPPKFVKLEASKVAKQGESIQLECKISG 3055  
QY 279 ---N--FI---L---SA-----GV-D--- 287  
Db 3056 SPEIKVSWFRNDSSELHESWKYNMFSINVALLITINEASAEADSGDYICEAHNGVGDASCST 3115  
QY 288 ---K---T-----T---IIW----- 293  
Db 3116 ALTAKPPVPTQKPSVPGALKGSDVILQCEISGTPPEVVMVKDRQVRNKKFKITSKH 3175  
QY 294 -D-----A---H-----T---G--E- 299  
Db 3176 FDTNLHLINLEASDVGEYHCATNEVGSCTCSVKFKEPPRFVKKLSDTSLIGDAVEL 3235  
QY 300 -A-----K-----Q----- 303  
Db 3236 RAIVEGFQPIVSVWLKDRGVIRESNTRISFDINIAITLQGSPEASNSGYICQIKNDA 3295  
QY 304 -----F-----P----- 306  
Db 3296 GMBECSAVLTVLEPARIIEKPEPMTVTGNPFALCVVTTGTPELSAKWFQKGRSLSADSK 3355  
QY 307 -H-S---A---P-A---L-----D-V--- 315  
Db 3356 HHITFINKVASLKIPCAEMSDKGLYSFEVKNVSKNCTVSVHVSDRIVPPSFIRKLKDV 3415  
QY 316 -----W-----QS-----N----- 319  
Db 3416 NAILGASVLECRVSGAPISVGVFQDGBEIVGPKQCSFSENVCTLNLSLEPSDTGI 3475  
QY 320 -----N-----TF----- 322  
Db 3476 YTCVAANVAGSDCSAVLTVQEPSPFEQTPDPSVEVLPGMSLTFTSVIRGTPPPKVKWFKG 3535  
QY 323 -----ASC----- 325  
Db 3536 SRELVPGESCNISLEDFVTELELPEVQPLESGDYSCLVTNDAAGSACTHLLFVKEPATFV 3595  
QY 326 -----S-----T-----D-----M----- 329  
Db 3596 KRLADFSVERGSPVLEATYGTTPPISVMSWKDEYLISQSERCSITWTEKSTILEILEST 3655  
QY 330 -----C-----I---HV-----CKL-G----- 337  
Db 3656 IEDYAQYSLIENEAGODICEALVSVLEPPYFTEPLEHVEAVIGEPAITQCKVDGTFEIR 3715

QY 338 -----Q-----D----- 339  
Db 3716 ISWYKHTKLRSAPAYKMOFPKNVASLVINKVDHSDVGEYSCKADNSVGAVASSAVLVIK 3775  
QY 340 -R---P-----I-KT-----F---Q-----G-----H--- 348  
Db 3776 ARKLPFPFARKLKVHETLGPVAFECRINGSPLQVSWYKDGVLKKDDANLOTSPFHVNV 3935  
QY 349 -----T-----N-----EV----- 352  
Db 3836 ATLQILQTDQSHIGQYNCASNPLGTASSAKLILSEHEVPPFDLKPVSVDLALGESGT 3995  
QY 353 -----NA-IK--W--D---PTGN-----L--L---A--SC--S--- 368  
Db 3896 FKCHVTGTAPIKITWAKDNREIRPGNYKMTLVENTATLTVLKVGGDAGQYTCYASNIA 3955  
QY 369 ---D-----DM-T-L-KI-----W---SMK-Q--- 380  
Db 3956 GKDSCSAQLGVQEPFRFIKKLEPSRIVKQDEFTRYECKIGGSPFIKVLWYKDET-EIQES 4014  
QY 381 -----D-----N-----C-----V-----H--- 385  
Db 4015 SKFRMSFVDSVAVLEMHNLNLSVEDSGDYTCFAHNAAGSASSSTSLKVKEPPIFRKXPHIE 4074  
QY 386 ---D---LQ-----AH----- 390  
Db 4075 TLKGADVHLECELQGTGPPFHVSWYKDRBLRSKKYKIMSENLFTSIHILNVDAADIGEY 4134  
QY 391 ---N-----KEI-Y--TI-----K 398  
Db 4135 QCKATNDVGSDDTCVGSIALKAPPRFVKKLSDISTVVGVEQLQTTIEGAEPISVWFKDK 4194  
QY 399 ---W-----S----- 400  
Db 4195 GEIVRESDNIMWISYSENIATLOFSRVEPANAGKYTCQIKNDAGMORCFATLSLEPATIV 4254  
QY 401 ---P---TG-----G--T-----N---N-- 408  
Db 4255 EXPESIKVTGDTCTLECTVAGTPELSTKWFQDKGKELTSDNKYKISFPFNKVSGLKINVA 4314  
QY 409 P-----N-----A-----N-L---M---L--- 415  
Db 4315 PSDSGVYSFEVQNPVKGDKSCTASLQVSDRTVPSPFTRKLKETNGLSGSSVWMECKYVGGSP 4374  
QY 416 -ASAS-P-----DS----- 422  
Db 4375 PISVSWFHEGNEISSGRKYQTTLTNTCALTVMNLEESDSGDYTCIATNMGSDCSAPL 4434  
QY 423 TVR-----L-----W----- 427  
Db 4435 TVREPPSFVQKPPMDVLTGTVNFTTSIVKGTTPPSVSWFKGSSELVPGDCNVSLEDSV 4494  
QY 428 -----DVD--R-G--ICI-----H----- 436  
Db 4495 AELELPDVTDSQSGYTCIVSNEAGKACTHLYIKAPAKFVKRLNDYSIEKGKPLILEG 4554  
QY 437 -----T----- 437  
Db 4555 TFGTTPPISVTWKNGINVTSPQRCONITTEKSPILEIPSTVEDAGQYNCYNIENASGKD 4614  
QY 438 -----L-----TK----- 440  
Db 4615 SCSAQIILLEPPYFVKQLEPVKVSVGDSASLQCLAGTPEIGVSWYKGTCLRPTTYKM 4674  
QY 441 H-----QE-----P----- 444  
Db 4675 HFRNNVATLVFNQVINDSGEYICKAENSVEGVSASTFLTVOEQKLPPSPSRQLRDVQET 4734  
QY 445 V-YSAF-----S-P-----DG-----R-----Y--- 456  
Db 4735 VGLPVVFDCAISGSEPIVSVWYKDRKPLKDSNVQTSFLDNTATLNI FKTDRSLAGQYSC 4794  
QY 457 LAS---GS-----FD-----KCVH-----I---W--- 469

```
Db 4795 TATNPICSSASSARLILTBCKNPPFFDIRLAPVDVAVGESADPEC-HVTGTQPIKVSMAK 4853
Qy 470 -----N-----T-----Q-T-----G-----A-----LVH 478
Db 4854 DSREIRSGGKYQISYLENSAHLTVLVKVDGDSGGYTCYAVNEVGKDSCTAQLNIKERLIP 4913
Qy 479 -----S-----Y-----L-----480
Db 4914 PSFTKRLSETBETEGNSFKLEGRVAGSQPIITVAMVKNKNIIEIOPTSNCEITFKNNLTVLQ 4973
Qy 481 -R-G--TG-----G-----I-----F-----E-V--C- 490
Db 4974 VRKAGNDAGLTYCKVNSDAGSALCTSSIVIKEPKPPVFDHLLTPVTVSEGEYVQLSCH 5033
Qy 491 -----W-----S-----NAAG---495
Db 5034 VQSEPIRQWLKAGREIKPSRCSFSPASGTAVLELRDVAKADSGDYVCKASNVAGSDT 5093
Qy 496 -----DK-----V-G-----AS-----D-----504
Db 5094 TKSQVTKDKPAVAPATKAAVDGRLFFVSEPSQIRVVEKTTATFFIAKVGGDPIPNVKT 5153
Qy 505 -----GSV-----CV-----L-----510
Db 5154 KGKWRQLNQGGRVFIHQKGDEAKLEIRDTTKTDSGLYRCVAFNEHGEIESNVNLQVDERK 5213
Qy 511 -----DLR--K 514
Db 5214 KOEKIEGDLRALMLK 5227

RESULT 4
T13931
projectin - fruit fly (Drosophila melanogaster) (fragment)
C:Species: Drosophila melanogaster
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T13931
R:Daley, J.; Southgate, R.; Ayme-Southgate, A.
J. Mol. Biol. 279, 201-210, 1998
A:Title: Structure of the Drosophila projectin protein: isoforms and implication for pro
A:Reference number: Z17815; MUID:98300339; PMID:9636710
A:Accession: T13931
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-6658 <DAL>
A:CROSS-references: UNIPROT:O76281; EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AAC2
C:Genetics:
A:Gene: projectin
A:CROSS-references: FlyBase:FBgn0005666
A:Map position: 4
A>Note: intron positions not resolved (incomplete sequence)
C:Keywords: muscle

Query Match 70.7%; Score 2594.2; DB 2; Length 6658;
Best Local Similarity 8.7%; Pred. No. 2.9e-52;
Matches 445; Conservative 55; Mismatches 11; Indels 4578; Gaps 397;

Qy 1 M-S-----IS-----S-----D-----E-VN-----9
Db 674 LPSSEIKNIPYNTKISIIETVRKHTGIYKIIIVNEHGGQDEATVEVNILAPPSPKRGPLD 733
Qy 10 -----F-----L-VY-----RY-----LOES 19
Db 734 VKDVTKDSCKFKWKPPEDDGGKPISAQVEKFDKQGRWVPLGRTSANDTEFDVKGLQE- 792
Qy 20 G----F-----S-----H-----S 24
Db 793 GHEYQFVKAINEGESDPLSDSDSIIAKNPYDAASKPGTPIVDYNEHWVKLWKEAPRS 852
Qy 25 ---A-----F-----T-----F-----G-----29
Db 853 DGGAPISGYIEKKDKFSPINWBEILSTNTSVPEATVEGLVEGNIYQFVRVAVNKAQFSDP 912
```

```
Qy 30 ---IESH-----I-----S- 35
Db 913 SDATPHLAKPRLNKPYNRDKMKPIKVRAGQPVKFDVDVKGEPAFLTWFLKETELTST 972
Qy 36 -----QS-----NING---A-----L-V- 44
Db 973 GOVRLNIDYNTKLTLLDTRKQSGQYKLAENINGVDBAVVEVILDKPSKPEGPLEVS 1032
Qy 45 -----P-----P-AAG-I-----SI-----Q- 58
Db 1033 DIHKEGCKLWKRKDDGGIPITGVIEKMDTATGKWVPAGSVDPPEKYDIEIKGLDPNHR 1092
Qy 59 Y-----V-----E-----A-----E-V-----64
Db 1093 YQFRVAVNEEGESEPLETESAITAKNPFVDSAPGCLPELEDWDEHHVKLWKEPPIRDGG 1152
Qy 65 S-I-N---E---D-G-----T-----L-----72
Db 1153 SPITNYIIEVMDKDPGEFVKAVETDSPVCKGVVKKLEEGQYKFRVRAVNKAGSPDSEQ 1212
Qy 73 -----F---D-----G-----R---P-IE-----S 80
Db 1213 TNHWAKPRFLKPHIDRVNLKPIVKTGLSISLDINIRGEPAPKVEWFFNNSSVTSDEHS 1272
Qy 81 -----L-----S--LI-----D-A-----V---87
Db 1273 VKIDNVYNTKFFVMRAQBSQSKYIIKATNEVGDEAELEVTVLGKPKGKPLQVNDI 1332
Qy 88 -----MPD-----V---V-----Q-----93
Db 1333 TKHSCKLWKEKDDGGSPIDYVEIEKLDPHGTQWLPCGKSTEPKAKVIGLHGKAYKFR 1392
Qy 94 -----T-----R-----Q-----Q 96
Db 1393 VRAVNKEGESELETEKPIIAKNPYDEPRPGKPEPTNWDKDFVDLAWPPKNDGGAPIQ 1452
Qy 97 ---Q-----A-----Y-R-----D-K-- 102
Db 1453 KTVIQMRDKSGRAWDSATVPDGKNGITVGVGEGHEYEFRIVAVNKAQSPSDVSKSV 1512
Qy 103 LA-----Q-----Q-H-----A-----108
Db 1513 IAKPRFLKPHIDRKNLQKKIMRSGQMLHIDALIKAEPPAKVTWYNTKBIKTSDHKIE 1572
Qy 109 -----A-----A-----A-----A---112
Db 1573 EDYKTTFMPKVRADRGIIYIVTAKNDSGDTVEVELEVLCCKPSKPKGLAVNSVTAETL 1632
Qy 113 -----A-----A-----A-----A---114
Db 1633 HLKWEKPEDDGGPIEQVLYVERMDTETGRWVPVLTTKTPEADVTGLTEGKEYLFRVAVN 1692
Qy 115 -----A-----AA-T-----N-----119
Db 1693 SEGESEPLVDIPTAKXNPFDAADTPGKPIQVDSGNHCDLKWRAPEDDGGASITGVIVE 1752
Qy 120 -----Q-----Q-----G-----S---AK-- 125
Db 1753 RKDPNTGKWKALETSTPDCARVNDLIAGNKYQFIRMAVNKAQSKPKSPSDQMTAKOR 1812
Qy 126 -----N-----N-----N-----N---129
Db 1813 FAPPKIDRTNIKDITTSKAGQHIRFDIKVSGEPPATKWHLNKALENDSNVDNIMESYR 1872
Qy 130 -----T-A-N-----G-E-----E-----135
Db 1873 TKLTVPISKRFHSGKYTLKAENESGRDEASFVIVLDKPGPEGLPLVTDVHKEGCKLKW 1932
Qy 136 N-----G---AH-TI-----A--NN---H---T---146
Db 1933 NAPLDDGGPLPDHDIYIEKMDVESGRWLPSPGRFKPSFAELNNLEPSHEYKFRVLAVNTEGE 1992
```

147 QY -----D--M--M--E-----VD-----152  
1993 Db SEPTIGQSUTAKNPFDEPGKGTPEAVDMDKOHDLVWRPPINDGGSPITGYVVEKREK 2052  
153 QY G-D-V--EI-----PS--N--KAV-----163  
2053 Db GTDKWIKGTETIPCLGCECKATVPTLNENCEYEFVRKAINAAGPGSPDASKPIITKPR 2112  
164 QY -----VL-----R-----G-----H-E-----169  
2113 Db KLAFTILDPTKNIRTYNFKSGEPIFLDINISGEPADVTWNQNNKSVQTTFSFHIENLPY 2172  
170 QY -S-----E-----V-F--I-----C-----175  
2173 Db NTKVINNNPERKOTGLYKISAHNFYQDQVEFQINIITKPKGPGPLEVSEVHKDCKLK 2232  
176 QY -----A-----W--N-----PVSD-----LLV-----185  
2293 Db ESEPLETGSIIAKDPFSVPTKPGVPEPTDWTANKVELAWPEPASDGGSPIOGYIVEVKD 2352  
186 QY -----S-----G-----SG-----D-S--TA-----193  
2353 Db KYSPLWEKALETNSPTTATVOGLIEGNEYQFRVVALKNLSEPSDPSKIIFAKKRYIA 2412  
194 QY -----R-----I-----W-----N-----L-198  
2413 Db PKIDRRNLRLITLSSGTALKLDANI TGQAPKPVKEWKLNVHLQSGKNVTIETPDYTKLV 2472  
199 QY -----S-E-----N-----S-----T--S--GP-----206  
2473 Db IRPQRTDSGYLVATNTSGKDSVLNVVITDKPSPNGPLQISDVHKEGCHLKWKPS 2532  
207 QY -----T--Q--L-----V-----L-211  
2533 Db HDGTPLEYFQIDKLEBETGCWIPSCRSSTAPQVDVTGLSPGNEYKFRVSAVNAEGSQPL 2592  
212 QY -----R-----H-----C-I-----215  
2593 Db VGDESIVARNPFDEPGKENLKATDWDKOHVDLAWTPPPVIDGGSPISCVIIERQDKYKWK 2652  
216 QY -R-----EG-----218  
2653 Db ERALDVPADOCKATI PDLVEGQYKFRVSAVNAAGTGEPSDSTPPIIAKARNKPIIDRS 2712  
219 QY -----GQ--D-V--P-----S-----N--K-----226  
2713 Db SLVEVRKAGOSFTFDCKVSGEPAPQTKMLKKKEVYKDNVKTVDYNTKLKVNSATR 2772  
227 QY -----D--VT-----S-----230  
2773 Db SDSGIYTVFAENANGESADVKVTVIDKPPAPNGPLKVDEINSECTLHWNPPDDGGQP 2832  
231 QY -----LD--W-----N-----SE-----236  
2833 Db IDNYVVKLDETTGRWMTAGETDGPVTALKVGGLTFCHYKFRVAKNRQGTSEPLTTAQ 2892  
237 QY -----GT-----L--L-----A-----TG-----243  
2893 Db AIIAKNPFDPVTPKGTPTIKDPKDFVLEWTRPEADGGSPITGYVVEKRDKSPDWEKC 2952  
244 QY -----S-----Y-----245  
2953 Db AEISDDITIAHVPDLIEGLKLEFRVAVNAKAGSPSDATETHVARPKNTPPKIDENFMS 3012  
246 QY -----D-----G-----F-----248  
3013 Db DIKIKAGNVFDPVPTGEBPLPSKDWTHEGNMIINTDRVKISNFDRTKIRILSATSDTG 3072  
249 QY -----AR-----I--W-----T--253

3073 Db VYTLTARNINGTDRHNKVKTILDAPSVPEPALNGDVSKNSIVLRWRPPKDDGGSEITHY 3132  
254 QY -----K-D-----G-----NL-----A-----S--T--261  
3133 Db VVEKMDNEARWVPVGDCTDTEIRADNLNLDHYSFRVRAVNKQOSQPLTTSQPIAKD 3192  
262 QY -L-----GO-----H-----K--G--PI-----F-----A-----271  
3193 Db PYSHDPKQPOQATDQWKEHFVDLEWSTPKRGDGGAPISYIIKRPFGOWERAUVLGDN 3252  
272 QY -----L-----K-----273  
3253 Db CKAHVPELTNGGEYEFVRVIAVNRGGSPDPSSTIICKRFLAPFDFKSLMLNDITVHAG 3312  
274 QY -----W--N--KK-----G--N--F--I--L--SA-----284  
3313 Db NGLGWTLPTEASPRPLITWLYNGKEIGSNSRSGSLGFQNELTPEIIVSSLSRSABGRYTLIL 3372  
285 QY -G-----V-----D--K--T-----T-----II-----292  
3373 Db KNEHGSFDSAHATVLDLRPSPPKPLDITKITRDGCHLTWNVPDDDDGGSPILHVIIEKMD 3432  
293 QY -----W-DA-----H--T-----GS-----AK-----301  
3433 Db LSRSTWSDAGMSTHIVDVTRLVHRKEYLFRVKAVNAIGESDPLEAVNTIIAKNEFDEPD 3492  
302 QY -----L-----Q-----302  
3493 Db APGLIITDWRDHDLOWAVPKSGGAPISEVIIQKKEKSGPYTNVHRVPSNKNTTII 3552  
303 QY -----Q--P-----P-----FH--307  
3553 Db PELTEGEYEFVRVIAVNAQOSEPSEPMIMRKYPLPKIITPLNEVRIRKGLIFHTD 3612  
308 QY -----307  
3613 Db IHFIGEPAPEATLNSNPLLSNDRSTITSIGHHSVVHTVNCQRSDSGIYHLLLRNSSGI 3672  
308 QY -----307  
3673 Db DSGFELVLDLRPGPEGPMEEITANSVTISWKPCKDNGGSEISSYVIEKRLDTHGGG 3732  
308 QY -----SA-----P-----A-----L--D-----V-----314  
3733 Db WVPVNVYSKYNHAVPRLLEGTMVELRVMAENLQGRSDPLTSDQPVVAKSQYTVPGAP 3792  
315 QY -----D-----W-Q--SN-----NT-----321  
3793 Db GKPELTDSDKNHITIKWKQPI SNGGSP IIGYDIERRDVTNGRWIKINGQVPTAEYQDDR 3852  
322 QY -----322  
3853 Db VTSNHQYQVRISAVNAAGKTSPTSAPFNARPLREKPRFYFDGLIGKRIKVRAGEPVNL 3912  
323 QY -----A-----SC-----325  
3913 Db NIPISGAPPTIEWKRGDLKLEEGKRI SYETNSERTLFRIDDSNRDRDSKGYTVTAANERG 3972  
326 QY -----S--T--DM-----329  
3973 Db KDTADIEVIVDVKPSPGSPISYTTETAPDHISLHWSPKDDGGSDITGYIIBTFEGVDD 4032  
330 QY -----C-----330  
4033 Db WKPVFGTCFNTNFTVKNLVEGKKYVFRIRAENIYGASEALEGKPVLVKSPFPDPPGAPSO 4092  
331 QY -I-----H-----V-----C-----334  
4093 Db TISATYPSANLEWAPPDCCGKPIITGYIVERRERREGGEWIKCNYPPTNTSYTVSNLRDG 4152  
335 QY -----K--L-----G-----Q-----DR-----P-----341

```
Db 4153 ARYEFRLVLAWEAGPHGSKSPDMPTAEHQYRDPPEPPKPDRIITRNGVTLSWRPRTD 4212
QY 342 --IK--T--F-- 345
Db 4213 GKSRIKGYVEMRPMKQDKWTNDIPINSTVTVPVPSLKEGEBYSFRVVAENEVGRSDPS 4272
QY 346 --Q--G-- 347
Db 4273 KPSQPTIEEQPNKPCWELGKVRDIVCRAGDDFSIHVPYLLAPPKPNAFWYNDNMLDDNN 4332
QY 348 --H-- 348
Db 4333 RVHKLHTDDAASVVVVKSRKRDGQYRLQKDTSGEDTATINVRVLDSPPTLRADVF 4392
QY 349 --N-E--T--N-E-- 351
Db 4393 SGDSLTYWNPDDGGAIONVYIEKKEARSTWSKVSFCTVLFVRINLVNKEYDF 4452
QY 352 -V--N--A--I-K--W--D--P-- 359
Db 4453 RVIAENKYGSDPANTSEPIARHPFDIPNTPGIPHGIDSTEDSIATWTKPKHGGSP 4512
QY 360 TG--NL--L--A--S-- 366
Db 4513 TGYIEKRLSDDKWTKAVHALCPDLSCKIPNLIAENAEYFRVAANAAGQAYSGL 4572
QY 367 --C--SD--DMT--LKI--W-- 376
Db 4573 IFCRRPPHAKITSLSIPDMTVIAGDEFRITVPHYASPRPTASWSLNGLEVIPGERIKF 4632
QY 377 --SM--KOD--N--CVH--D--L--Q-- 388
Db 4633 DSNDAVSMYKNSAKRDEGTSYITITNKNKSDTASC-HVTVDRLPLPQGPLNAVDTIP 4691
QY 389 --A--A--H-- 390
Db 4692 DTCTLAWKTLDDGGSPITVYVVEKLDNSGWSWKISSFVRNTHYDVMGLEPHYKYNFRVR 4751
QY 391 --N--K--E--E--IYT-- 396
Db 4752 AENQVGLSDPLIEPMWAKHFTVPDEPGQPKVIDWDSGNVTLIWTRPLSGGGRIOQY 4811
QY 397 -IK--W--KTYQYLNLSNGSEYEFRIKAKNAAGLSKSPSPSLRFLK 4871
Db 4812 QIEYRDILNDSSWNAVYIIKGTQYLNLSNGSEYEFRIKAKNAAGLSKSPSPSLRFLK 4871
QY 403 -G--PG--T--N--NP-- 410
Db 4872 KGKFTVPSPGAPQVTRVGKNTYDLKWEKPLRDGGSRIITGYIERRDIGGAVVVKCNDYN 4931
QY 411 --ANL--M--LA--SA--S-- 419
Db 4932 VLDTEYTNVNLIEMGDYFRFRVAVNSAGRSFSLCTMPIKVEVLGGKPKDWTIRLQDKV 4991
QY 420 --F--D--S--TVRLW--DV--D--R-- 431
Db 4992 APFGKDYTLQCAASGRKPSPTAR-WLRNGKEIQMNGGEMTCDSKDGVRFLHISNVQTGGD 5050
QY 432 --QEP--G--I--C-- 434
Db 5051 DYTCEAMNSLGFVNTSGYLKIGSPIINRCPSBLKLPEDGNSKIKIPYSGDQPLTVILKKN 5110
QY 435 --I--H--T--LT--K--H-- 441
Db 5111 NEVISDSNDTHVKVNIFFDYVAIYANIVKSDGGPYQIEFTNESAATGEFYVHTGMP 5170
QY 442 --QEP--V--Y--S--V--AP-- 450
Db 5171 SAPTGMGILYINKNSCMLNWRPPYDGLKSHYVYTERKDVSSPHWITVSTCKDTAFN 5230
QY 451 -----SP----- 452
Db 5231 VQGLIENQBYIFRVMAVNMNGMPPLEGLNPIRAKODIPDPPSPFPAVLKSLRSEETLCNPE 5290
```

```
QY 453 -----D-----GR-Y----- 456
Db 5291 WKPPSDGGAHQYQWIDKREVGSKHMGVNTATICAANQINCINLEGRQYEFRIFAQNV 5350
QY 457 --L--A--S--GSF--D-K--CV--H-- 469
Db 5351 AGLSTAIVRFTSSQDNRSSTGGFASIDCETATHANCIONHNAQFTCTIPGVPKPTISWYKG 5410
QY 470 -----TQ--T-----GA----- 475
Db 5411 AREISNGAGITCTOKVTSPLPKYDVFGDEADYVCRVAVNKAGAKSTRATLAIMTAPKLV 5470
QY 476 -----L--VH-----S--Y-----R- 481
Db 5471 PPRFRDTAYFDKGVNVVVIKPTGLPKPRIHVHVGLENIESGGHYTVVEKHERHAVLIIRD 5530
QY 482 G--TG-----G--I-----F-----V--C 490
Db 5531 GSHLDGSPYRITAENELGSDTAIQVQISDRDPDPFRFPLIESIGTESLSLWKPVPWDGC 5590
QY 491 -----W--N--A-----A--G--D-- 496
Db 5591 SDITVYVERRDDPLSSWIRVGNTRFTSMVSLTPGKEYDFRIFADNVYGRSDASDTST 5650
QY 497 -----K--V-----G--A-----S-AS- 503
Db 5651 LIKTESVKKKPIERKWEIDANGKRLGKADGPKVDYSDYVFDIYKFKVPQVPEISQQSV 5710
QY 504 -----D--GS-----V--C-----VL-----DL--RK 514
Db 5711 YDRYDILEIGTGAFGVHRCRSTGNIFAAKFPVSHSVSKOLIRRE 5759
```

## RESULT 5

T30192

probable peptide synthetase - Aureobasidium pullulans

C:Species: Aureobasidium pullulans

C&gt;Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 09-Jul-2004

C:Accession: T30192

R:Peery, R.B.; Thorneswell, S.J.; Tobin, M.B.; Skatrud, P.L.

submitted to the EMBL Data Library, January 1997

A:Description: Discovery of an MDR-like gene adjacent to a peptide synthetase in Aureobasidium

A:Reference number: Z20767

A:Accession: T30192

A&gt;Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-10797 &lt;PEE&gt;

A:Cross-references: UNIPROT:O94116; EMBL:U85909; NID:g4099310; PID:g4099313; PIDN:AAD0058

C:Genetics:

A:Introns: 2078/3; 4142/3; 6075/3; 7963/3; 7985/2

C:Keywords: carrier protein

F:1618-1688/Domain: acyl carrier protein homology &lt;ACP1&gt;

F:3682-3752/Domain: acyl carrier protein homology &lt;ACP2&gt;

F:5615-5685/Domain: acyl carrier protein homology &lt;ACP3&gt;

F:7503-7573/Domain: acyl carrier protein homology &lt;ACP4&gt;

F:9683-9752/Domain: acyl carrier protein homology &lt;ACP5&gt;

Query Match

Best Local Similarity 70.7%; Score 2593.7; DB 2; Length 10797;

Matches 461; Conservative 40; Mismatches 8; Indels 4923; Gaps 386;

```
QY 1 M-----SI--S-----S-----D----- 6
Db 4302 MLNTQVIOQFGSGIKSGILYGMVGTPEAAIHTCTQPGFVDLPAGTIGIPDITVSCFIV 4361
QY 7 -----E-V-----N-----FLV-----YR--- 14
Db 4362 KPESTKHSQLEILPICEIGELVIGGHQLADGYLNRESEQTAAAF-VTHPKFGLYRTGD 4420
QY 15 -----Y-----L-Q-----E-----SG-----F--- 21
Db 4421 KARLHRNGTLECYGRISGQVKLRQORVELGIEHAASKAGGCHAVIASVISGLLVLCI 4480
```

QY 22 --SH-----SA-----P-----T-P-----28  
Db 4481 GDPRVSSKDIKACQKWLPAWMPISDIVLDDPPYLPSCVKDKRLETDYNSNTAHOVS 4540  
QY 29 G-----IES-----H-----I--SQ-----36  
Db 4541 GSSDLSENAREITRIIESVLGVSDHSTDLASAAGLDSLRAIQVASQLRRQGCADLGALEL 4600  
QY 37 -----SNIN-----G-----41  
Db 4601 LSVSNVLALDELVRAKADESNINDNDSEKWKQTVHELRSVERDFESKAFVSGIEDVLP 4660  
QY 42 -----A-LV-----P-P-----A-----AL-----IS-51  
Db 4661 TPLQDAMLVETAKPQAYCNELRLTVSPKIPVERVROALFALAQHRTALRSGMPSGVSH 4720  
QY 52 -----II-----Q-K-----G-----L-QY-----VEAE--V-----64  
Db 4721 CAVTQVIWKTIVTSQFAHVKSFTTGMVNTNRETLRLPHFYKCSGAEBILVAIHVALY 4780  
QY 65 --S--I-----NE-----D-G-70  
Db 4781 DQMSVEILEDLETLQONERTPERPSFGAVNKFNLRSBEDQTSHLDPWGEYLSVTPGR 4840  
QY 71 -----TL-----F--D-74  
Db 4841 LPNLSPKMPPPQLOSIOHTIEMDMETLROAAHSYSCSAHVFFQAAAYAILLGYMGTEDT 4900  
QY 75 -----GR-P-----LS-L--I-----DA-----VM-----88  
Db 4901 VFGTVSGRTLPIVEIESMVGPLLSTLPTINTLESKRFSDSLRLQEDNRKIMRHSMTS 4960  
QY 89 -----P-----D-V-V-Q-T-----R-----OO-97  
Db 4961 LADIKKACGNPEAVFDSIFVQWETARPDPARAQTLNLNVEAHDYLEFNLTLBLEPTQQ 5020  
QY 98 --A-YR-----D--KL--A-----Q-----O-106  
Db 5021 VKTKATYQSSLLPIQHVKTLLQOLDALVKIIVARPETHMNEISDQLPISVLVANSPEQS 5080  
QY 107 -----HA-----108  
Db 5081 FVYKAGLSIVENHALNNSGGLAVFAHDIRECTSRMESITYGELNTRANQANLYISQ 5140  
QY 109 -----AA-----112  
Db 5141 AKRDELICVMEKSVSLYLSILAAVKAGCYLPLVPETPAARIRQILAEADVFKCLTDS 5200  
QY 113 --A--A-----A-----A-A-----T---118  
Db 5201 MAPVIADVSRCHIMNVDTDCSAQSGTGPOLDPKPTDIAVAVFTSGTGPKGVLVTQEN 5260  
QY 119 --N-----Q-Q-----G-----SA-----K-125  
Db 5261 ILSNLEVLKIPVPEGSRLLOACNOAFDVSVFPIFTWTGMCLCSASKDMVFRDFEKA 5320  
QY 126 -N-----GE-----128  
Db 5321 INELEITHLSLTPVAALTDPAHPRVKFLVTAGEAVTHHVHGAWAGKGLYQGYGPSETT 5380  
QY 129 N--T-----N-GE--EN-----GA-----138  
Db 5381 NICTVNSAVESDHVINNIAPAPENTSFAVLVTQGDDEFQVLPLGGLGELCGFGQQVFRGYQN 5440  
QY 139 -----H-----TT-----141  
Db 5441 MPELTSKINHPNYGRIVRSGDLRLLPDGTILIQRTDDQKIRGQRIELGESCILL 5500  
QY 142 -----AN-----N-----H-----145  
Db 5501 QPFSVQNCALIEVKTADKRLMAFWIPSGVSKDSYSILOPKMLBILIKSIYAHILADNLP 5560

QY 146 -----T-DM-----M--E--150  
Db 5561 AYVVPDALVPVSAIPOTSQKIDKRRLASDGGALTVEDLNAYSRGADDDTSELSATEQQ 5620  
QY 151 -----VD-----G-D-V-----E-----I--157  
Db 5621 LASALADTLQMSQTSIGRSTSPFALGLDSVAIRLATNLNRKEYGYSIDVSQLKRRTIAR 5680  
QY 158 --P-----S--NK-----AV-----VL-----165  
Db 5681 LAPLLGGESSKOTNEPVTADCEAAAVGSLHDESQVLSQHEHGQTVSQVLPCTPQLOEAMLS 5740  
QY 166 -R--G-----H-----E-----S-E--V-PI-----174  
Db 5741 ARDTSGSSAYRNKTLFSLHSGVDKLCACWEVMLQRHDIILRTIFLSTEDSRFPFQVAVLSQ 5800  
QY 175 -----C-----A-----W-----177  
Db 5801 WTLPMQECDDIPDQLSTLLDSAKAGGDSIVDHPMPWKIQVYRSESTVYLLDMHHALYDA 5860  
QY 178 --N-----PVS-----181  
Db 5861 NAMSNNLLYEVOQLYKQDQSLSAFVSPKPLNFMISTSVBEADALFRDQLREFVPKPKRTD 5920  
QY 182 -----181  
Db 5921 VKSGFGTITGRNLNYSKPVETFLSKHSTTMLSLTQAMWMTLAASQSYSDVCGNVVSGR 5980  
QY 182 -----DL-----183  
Db 5981 SVPVDGIESLVAFCFNTPVRVDLSKHSNGLVKALQRVNDISLPYQLTPLRRIQAQAG 6040  
QY 184 -----LV-----S-----G-SG-D--ST-----192  
Db 6041 TNGKRLFDLSVLQQDPTDLDLSAIWRLEGESGMDMTYTGSTGLPKAVCLSHRAVTSLL 6100  
QY 193 A-----R-----I--W-----N---197  
Db 6101 AHDRTIPSRFLQFASPTFDVSVEIFEIPFPWYRGATLVSVERNLLGDLPGTITSNIDA 6160  
QY 198 -L-----S--EN-----S-----T-----SG-----205  
Db 6161 ABELTPSAASLVRRHRENVPTLRALLTIGEMLNTQVIOQPGSIGKSGILYGMYPTEAAI 6220  
QY 206 -----PT-----19-209  
Db 6221 HCTLOPGFVGLPAGTIGIPLDTVSCFIVKPTSTKHAQLEILPIGEIGELVIGHQLA 6280  
QY 210 -----V-----L-----R-H--C--I-----R-----216  
Db 6281 DGYLNREEQTRAAFTVTHPKFGLYRTGDKARLHRNGTLECYGRISSGQVKLRQORVELGE 6340  
QY 217 -B-----GG-----Q-----D-V-----222  
Db 6341 IEHAASKAGGCHAVIASVISGLLVLCIGDPHRVSSKDIKSAQKWLPAWMPISDIVLDD 6400  
QY 223 -----P-----SN-----KD-----V-----T-----229  
Db 6401 DFPYLPSCVKDKRLETDYNSNTAHOVSGSSDLSENAREITRIIESVLGVSDHSTDLA 6460  
QY 230 -----SL-----D--W---233  
Db 6461 AGLDSLRAIQVASQLRRQGCADLGALELLSVSNVLALDELVRAKADESNINDNDSEKWKQ 6520  
QY 234 -----N-----S-----235  
Db 6521 TVHELRSVERDFESKAFVSGIEDVLPCTPLQDAMLVETAKPQAYCNELRLTVSPKIPV 6580  
QY 236 E-----G-----TL-----LA-----TG-----243  
Db 6581 ERVROALFALAQHRTALRSGFMPSGVSHCAVYQVIWKTIVTSQFAHVKSFTTGMVNTNRE 6640  
QY 244 -----S-----YD-----G-----247

Db	6641	TLRLPLHFOYKCSGABAEILVAIHHALYDQWSEVEILEDLETLQNERTPERPFGAVNK	6700
Qy	248	-PA-R- -I-W- -TKD- -G-NL- -	258
Db	6701	FFNLRSSEDTSHLDWFGEYLS-DVTPGRPLNLSPKMMPQPOLOSIOHTIEMDMETIROA	6759
Qy	259	- -A- -S-TL- -G- -	263
Db	6760	AHSYSCSAHVFFQAAVAILLGFYMGTEDTVFGVFSGRITLPIVEIESMVGPLLSTLPTRI	6819
Qy	264	-Q- -H- -K-G-P- -IFALKW- -	274
Db	6820	NTLESKFSVLSRLOEDNRKIMRHSMTSLADIKKACGFNPGFAVDSIFV-WQETARP	6877
Qy	275	- -N- -	276
Db	6878	DARAQTLLNLVEAHYLEFNLTLELEPTQGVKTAKTYQSSLLPLQHVKTLLQQLDALVK	6937
Qy	277	- -K-G- -N- -F- -	280
Db	6938	IIVARETHMNEISDQLPISVLSVANSEPOSFYKAGLSLVENHALNNSGGLALVFAHD	6997
Qy	281	- -I- -LS- -AG- 285	
Db	6998	IREGTSRMESLTYGELNTRANQLANYLISQAKRDELICVMEKSVSLYLSILAALKVAGC	7057
Qy	286	- -VD- -	287
Db	7058	GYPPLVPETPAARIQILAEADVFKCLTDSSMAPVIADVSRCHIMNVDTTDCSAQSGTGP	7117
Qy	288	- -K- -I- -	291
Db	7118	QLDFKFTDIAYAVFTSGTCKPKGVLTQENTILSNLEVLKIPVPEGSELLOACNQAFD	7177
Qy	292	- -I-W- -D-AH- -	296
Db	7178	VSVEIFFFTWYTGMLCSASKDVMFRDFEKAINLEITHLSLPTVAALTDPAHIPRVKF	7237
Qy	297	- -T-GEA- -K-Q- -	302
Db	7238	LVTAGEAVTHVHGAWAGRLYGYGPSETTICTVNSAVESDHVINNIGPAPENTSAPV	7297
Qy	303	- -Q-F-P- -F- -H- -SA- -P 310	
Db	7298	LTQGDFFQLVPLGELGELCFGGQVFRGYQNMPELTESKIINHPNYRISGDLGRLLP	7357
Qy	311	- -AL- -D- -V- -D- -W- -	316
Db	7358	DGTILIQRTDDQKIRGQRIELGEISGCLLOPPSVQNCALIEVIKTADKERLMAFWIPSG	7417
Qy	317	- -Q- -S- -N- -	319
Db	7418	YSKDSYSILQPKNLBEIKSIYAHADNLPAWVPDALVPVSAIPQTSQKIDKERRLAS	7477
Qy	320	- -N- -T-FA- -	323
Db	7478	DGSALTVDNLNAYRGADDDTSELSATEQQLASALADTLQMSQTSIGRSSTFFALGLDS	7537
Qy	324	- -S-C- -T- -DM- -C- 330	
Db	7598	HDESVVSQLHEHQTVSQVLPCTPQLOEAMLSARDTSGSSAYRNKTLFSLHGSVDKLKACW	7657
Qy	331	- -I- -H- -V- -C- -K- -L- -G- -	337
Db	7658	EVMLQRHDLIRTFLESTEDSRPFVQAVLSQWTLPMQECDDIPDQLSTLLDSAKAGGDSI	7717
Qy	338	- -Q- -DR- -PI- -K- -	343
Db	7718	VDHSPWKIQVRSSESTVYLLLDMMHLYADANAMSNLLYEVEQLYKQDLSAPVSPKFL	7777
Qy	344	- -T- -F-Q- -G- -H- -	348
Db	7778	NFMI STSVBEADALFRDQLREFVPKPKRTDVKSGGTTTGRNLNYSKQWVETLSKHSTT	7837
Qy	349	- -T- -N- -E- -V- -	352
Db	7838	MLSLTQAMMKTLAASQSYSDVCCGVNVSGRSPVDGIESLVAPCFNTIPVRVDSLCHRS	7897
Qy	353	- -A- -I- -K- -W- -	357
Db	7898	NLGLVKALQVNIIDSLPYQLTPLRRIQAOAGTNGKLFDSLVLLOQDITDLSA1WRLEG	7957
Qy	358	- -D- -PT- -GNL- -L- -ASC- -SD 369	
Db	7958	ESGVMDFPCIVELAPTNESYTLSLHFNRSYLDDEVVSNLHQACLSAFASCIIRYPSSDVSD	8017
Qy	370	- -DM- -TLK- -I- -WS-MK-Q- -	380
Db	8018	FIDFDADLVAGVLKPDTHMQPVEAAKTRSEKSGSGDESHSPLEQIRAAVSAVSSAP	8077
Qy	381	- -D- -N- -C- -V 384	
Db	8078	EDRIRRDTTIYKGLDLSISAIQLANRLRKDGLLVOASDVWESPSCSELASAVQSRQTPV	8137
Qy	385	- -H- -DLQ- -A- -H- -N- -	391
Db	8138	LBERGDFEGFKYRGGAALQSHRIATEKVASVRPCTPLOSGLMSEYTHSDGHQYNNHTF	8197
Qy	392	- -K- -E- -I- -YT- -	397
Db	8198	YAEADIDSSKLQSAWSKVLEQHELLRTGCTGDDHHEHPFVMLTYTTFDVIDLEIQASSR	8257
Qy	398	- -K- -WS- -	400
Db	8258	EGSVVEYSEKASESVKNNLHLPWRWSLLGVGRCLOFSAHHAIFAESLRLIMTDLQ	8317
Qy	401	- -PT- -G- -P- -	404
Db	8318	SALSNYVPTLITIDGALHILNSQADVESORTFWSQKLSGAPVTRFPNMTFVRISDTE	8377
Qy	405	- -G- -	405
Db	8378	AANVELVLYKSKLEARCOELGVSMQSVGOAARLLSAYTGESQVTFGVLSGRITSPA	8437
Qy	406	- -TN- -N- -P- -NA- -N- -	412
Db	8438	TADAAFPCCITTLTVSTNTAVDDSQFLKDLMSYNATIQHQFTPLTNIRNVAESTSEALFD	8497
Qy	413	- -L- -M- -LA-S- -A-S- -	419
Db	8498	SUFVYQRMNDYVDSSWKIIREKASVELAVSMEALSEDGLGLRLTIDPAQVPEHQK	8557
Qy	420	- -F-D- -S- -T- -V- -R- -	425
Db	8558	IMLQOMEVMIAGLLKFEDAINTSVMSIIPPKDPITATDFKYLHEMTEASVKSVDRIAME	8617
Qy	426	- -L- -W- -D- -VDRG- -	432
Db	8618	FVDALEDGQISSRHWTYRQDDEANKIAHLIDRGVKPGDIIATSPDKCEASFAFYGIL	8677
Qy	433	- -I- -	433
Db	8678	KAGCAFAIDPTAPARKAFILEDNSNARVLLTSDSIRSELRELTQCDIIDLINFENKNEL	8737
Qy	434	- -C- -I- -	435
Db	8738	STSSVPVSGLAPSSVSVLYTSGTTGTPKGCETIHDNAVQLVMSFKRLFKGRWTDSERL	8797
Qy	436	- -H- -T- -LTK- -	440
Db	8798	QFASYHFDVSVLEQFWMTWVGMRVLCAPRDLILEDIAGFLDTMQITHLDTLPSLGRLLDP	8857

[illegible]

## RESULT 6

I38344  
titin, cardiac muscle [validated] - human  
N:Alternate names: connectin  
N:Contains: serine/threonine-specific protein kinase (BC 2.7.1.-)  
C:Species: Homo sapiens (man)  
C:Date: 12-Aug-1996 #sequence revision 12-Aug-1996 #text change 09-Jul-2004  
C:Accession: I38344; I38345; S20898; S20897; S20899; S63665; S37393  
R:Labelit, S.; Kolmerer, B.  
Science 270, 293-296, 1995  
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.



,21900,21935,22295,22627,22897,23024,23318,23883,24012,24177,24290,24447,24642,248  
F;26171,26176,26184,26190/Binding site: phosphate (Ser) (covalent) #status experimental

Query Match 70.7%; Score 2592.2; DB 1; Length 26926;

Best Local Similarity 9.0%; Pred. No. 5.8e-51; Indels 4368; Gaps 365;  
Matches 437; Conservative 58; Mismatches 12;

QY	1	M-----SISD-----EV--N-----F-----L-----	11
DB	20024	MRFNTEITAENLTINLKESVTADAGRYEITAANSSTTKAFINIVLDRPGPTGPPVIS	20083
QY	12	-----V-----Y-----R-----Y-----L-----OE-----S-----	20
DB	20084	DITEESTVLKWBPPKYGDSQVNTYLLKRETSSTAVTVEVSATVARTMMKWLTTGEY	20143
QY	21	-----F-----F-----F-----F-----F-----F-----S-----H-----	25
DB	20144	QPRIKAENRFGISDHIDSACVTVKLPYTPPGPSTPMVNTVRESITVGMHPEVSVNGGSA	20203
QY	26	-----F-----F-----F-----F-----F-----F-----GI-----BSH-----	33
DB	20204	VVGYHLEMDRNSILWQANKLVIRTHFKVTIISAGLIYEFVVAENAGVCKPSHPSE	20263
QY	34	-----S-----S-----S-----S-----S-----S-----S-----S-----	33
DB	20264	PVLADACEPPRNVRITDISKNSVLSWOOPAFDGGSKITGVIVERRDLPGRWTKASPT	20323
QY	34	-----IS-----O-----S-----N-----I-----N-----	40
DB	20324	NVTETQTTISGLTQNSQYEFVRVARNAGVISNPSEVWGPICTIDSYGGVIDLPLEYTE	20383
QY	41	-----G-----G-----G-----G-----G-----G-----ALV-----	44
DB	20384	VVKYRAGTSVKLRAGISGRPAFTIEHWKDDKELOTNALVCVNTTDLASILIKDADRLNS	20443
QY	45	-----P-----P-----P-----P-----P-----P-----PPA-----A-----	48
DB	20444	GCVELKLRNMAASATIRVQILDKPGPGPIEFKVTVAETKILLWRPPADGGAKITH	20503
QY	49	LI-----S-----S-----S-----S-----S-----S-----I-----	53
DB	20504	YIVEKRETSRVVMSVMSVSEHLEBECIITTKIIRKNEYIFRVAWNKYGIGEPLESDSVAK	20563
QY	54	-----S-----S-----S-----S-----S-----S-----S-----S-----	53
DB	20564	NAPVTPPGPIPEVTKITKNSMTVWSRPIADGGSDISGFLEKRDKXSLGFWKVLKETI	20623
QY	54	-----QK-----GL-----V-----V-----V-----V-----E-----	61
DB	20624	RDRQKVTGLTENSQYRVCVNAAGQGPFSEPFYKAADPIDPPGPPAKIRIADSTK	20683
QY	62	-----A-----A-----A-----A-----A-----A-----E-----	63
DB	20684	SSITLWSKPVYDGGSAVTGYVVEIRQGBEEHWTVTSTGEVTEYVYVSNLKPQVNYF	20743
QY	64	-----VS-----I-----NE-----D-----D-----D-----D-----G-----TL-----F-----	73
DB	20744	RUSAVNACQGFIEIENEPVQAKDILEAPEIDLDVALRTSVIAKAGEDVQVLIPFKGRPP	20803
QY	74	-----D-----G-----R-----R-----R-----R-----P-----P-----P-----	79
DB	20804	PTVTRKDEKLGSDARYSIENTDSSLLTIPOVTRNDTGKILTIENGVEPKSSTVS	20863
QY	80	-----S-----S-----S-----S-----S-----S-----L-----	81
DB	20864	KVLDTPAACQKLVKHSRGVTLWDPPLIDGSPIIINVIKRDATKRTKTSVSHKGS	20923
QY	82	-----S-----LID-----L-----L-----L-----L-----A-----V-----	88
DB	20924	STSFKLIDUSEKTPPFVFRVAENEGIGEPCTTEPVPKAAEVPAPIRDLMSKDKSTSVI	20983
QY	89	-----PD-----PD-----PD-----PD-----PD-----PD-----Q-----QA-----YR-----	100

DB	20984	LSWTKPDFDGGSVITEYVYVERKKGQETWSHAGISKTCTBIEVSQLEQSVLEFRVFAKNE	21043
QY	101	-----D-----D-----K-----L-----L-----L-----AO-----Q-----H-----	107
DB	21044	KGLSDPVTIGPITVKEIITPEVDLSDIPCAQVTVRIGHNVHLELPYKPKPKPSISWLKD	21103
QY	108	-----A-----A-----A-----A-----A-----A-----A-----A-----	110
DB	21104	GLPLKSESEFVRFSTENKITLSIKNAKKEHGKGYTIVLONAVCRIAVPITVITLGPSPK	21163
QY	111	-----A-----A-----A-----A-----A-----A-----A-----A-----	112
DB	21164	KGPIRDEIKADSVILSWDPEDNGGEITCYISIEKRETSQTNMKMVCSSVARTTTPKVPN	21223
QY	113	-----A-----A-----A-----A-----A-----A-----A-----A-----	117
DB	21224	LVKDABYQFVRBAENRYGVSOPLVSSIIVAKHQFRIPIPGPKPVIYVNTSDGMSLTWADP	21283
QY	118	-----T-----T-----N-----Q-----Q-----Q-----Q-----Q-----	123
DB	21284	VYDGGSEVTGPHVEKKERNLSILWQVNTSPISGREYRATGLVEGLDYQFVVAENSAGLS	21343
QY	124	-----A-----A-----A-----A-----A-----A-----K-----N-----	128
DB	21344	SPSDPSKFTLAVSPVDPGCTPDYIDVTRITITLKNPPLRDGGSKIYGVYSIEKROGNERW	21403
QY	129	-----N-----N-----T-----T-----T-----T-----A-----A-----	132
DB	21404	VRCNFTDSECOYTVTGLSPGDRYEFRIARNAVGTISPPSSOGSIIMTRDENVPPIVEF	21463
QY	133	-----GE-----GE-----GE-----GE-----GE-----GE-----E-----N-----	137
DB	21464	GPEYFDGLIISGESIRIKALVQGRPVPRVTFKQGVIEKEMNEITNVLSGSTSLFVRD	21523
QY	138	A-----H-----TI-----A-----N-----N-----N-----N-----	144
DB	21524	ATRDHRGVYTVAKNAGSAAKAEIKVKQDTPGVVGPPIRFTNITGEKMTLWMDAPLNG	21583
QY	145	-----H-----H-----T-----T-----D-----D-----D-----D-----	153
DB	21584	CAPITHYIEKRETSRLAWALLIEDCEAQSYTAIKLINGNEYQFRVSAVNFVGVGRPLDS	21643
QY	154	D-----V-----EI-----PSN-----PSN-----PSN-----PSN-----	163
DB	21644	DPVVAQIYQTVDPAGIPEPSNITGNSITLTWARPESDGSIEIQVYLERREKKSIRWK	21703
QY	164	VL-----R-----R-----G-----H-----E-----E-----S-----S-----	171
DB	21704	VISKRPISETRPKVTGLTEGNEYEFHVAENAGVGPASGISRLIKCREPVPNPPGPPTVV	21763
QY	172	-----V-----V-----V-----V-----V-----V-----I-----C-----	183
DB	21764	KVTDSTKTVSLEWSPVDPGDMBIEIYIEMCKTDLGDMHKVNAEACVKTRYTVTDLOA	21823
QY	184	-----L-----L-----VS-----VS-----VS-----VS-----TA-----TA-----	195
DB	21824	GEEYKFRVSAINGACKGDSCEVTGTIKAVDRLTAPELDIDANFKQTHVVRAGASIELFTA	21883
QY	196	-----W-----W-----W-----W-----W-----W-----N-----N-----	202
DB	21884	YQGRPTPAVWSKPDNSLSLRADIIHTDSFTLTVENCNRNDAGKYTLTVENNSGSKSIT	21943
QY	203	T-----S-----S-----GP-----GP-----GP-----GP-----L-----L-----	213
DB	21944	FVVKVLDTPGPPGPIITFKDVTGRSATLWMDAPLDDGGARHIIHYVVEKREASRRSQVISE	22003
QY	214	-C-----I-----I-----I-----I-----I-----I-----E-----G-----	222
DB	22004	KCTROIKVNDLAEGVPPYFRVSAVNEVGVGPEYEMPEPIVATEQPAPRRLLDVDTSKS	22063
QY	223	-----PSNKD-----PSNKD-----PSNKD-----PSNKD-----PSNKD-----	233
DB	22064	SAVLAWLKP-----DHDGGSRIITGYLEMRQKSDLWVEAGHTKOLTFTTVERLVEKTEYFR	22120

QY 234 ---N---SE---G-T--LL--A---T---G--- 243  
Db 22121 VKAKNDAGYSEPREAFSSVWKEPQTEPTADLTGNTQNLITCKAGSPFTIDVPSGRAP 22180  
QY 244 --SY---D-G--F---A--- 249  
Db 22181 KVTWLEEMRLKETDRVSITTKDRTLLVVKDSMRGDSGRYFLTLENTAGVKTFSTVVV 22240  
QY 250 --R---I---W---TKDG---NL---ASTLG--- 263  
Db 22241 IGRPGVTGPLEVSSVSAASCVLWSGPEKDGTEITNIVKRES--GTTAWQLVNSSV 22298  
QY 264 ---Q---H--K---G---PI---F--- 270  
Db 22299 KRTQIKVTHLTMYEVSFRVSSNRFVSKPLESAPIIAEHFPVPPSAPTRPEVHVHSAN 22358  
QY 271 A---L---K---W---NK--- 276  
Db 22359 AMSIRWEEPYHDGSKLIYGVWEKERNILWKENKVPCLCNCYKVTGLVEGLEVQFRT 22418  
QY 277 --- 276  
Db 22419 YALNAAVSKASEARPIAQNPDVADGRPEVTDVTRSTVSLIWSAPAYDGGSKVGYII 22478  
QY 277 --K---G---NF--I--LS---AGV--- 286  
Db 22479 ERKPVSEVGDGRWLKCNYYTIVSDNFTVTALSGDTYEFRLAKNAGVISKGSSESTGPV 22538  
QY 287 --- 288  
Db 22539 TCDEYAPPKAELDARLHGLDVTIRAGSDVLDAAVGKPEPKI IWTGDKELDLCEKVS 22598  
QY 289 --T---T--II--- 292  
Db 22599 LQYTKRATAVIFCDRSDSKYTLTKVNASGTAVSMVVKLDSPGCGKLTVSRTQ 22658  
QY 293 ---W---D--A--H---T---GE--- 299  
Db 22659 KCTLAWSLPQDGGAEITHIIVERRETSRLNWVIVEGECPTLSYVTRLIKNEYIFRVR 22718  
QY 300 ---A---K---Q--- 302  
Db 22719 AVNKGVPVSEPIVARNSFTIPSPGPIEVEGTGKEHII IOWTKPESDGGNEISNYL 22778  
QY 303 --- 306  
Db 22779 VDKREESLRWTRVKNKYVYVDTRLKVTSLMEGCDYQFRVTVAVNAAGNSEPSSNFISC 22838  
QY 307 ---HS--A--PALD---V---D-W-Q--SN-- 319  
Db 22839 REPSYTPGPPSAPRVDTTKHSISLAWTKEMWDGTDIVGVLEMQEKDQDQYRVHTNA 22898  
QY 320 ---NT---F--- 322  
Db 22899 TIRNTEFTVPLKMGOKYFRVAANVKNMSEYSESTAEIPEVRIEIPDLELADLKK 22958  
QY 323 --- 324  
Db 22959 VTIRAGASRLMVSGRPPPVITWMSQIGDILASRAIIDTTESYSLIIVDKVNRVYDAGKY 23018  
QY 325 --- 328  
Db 23019 TIEAENSGKKSATVLVKVYDTPGCPSPVKVKEVSRDSVTITWEIPTIDGAPINNYIVE 23078  
QY 329 ---M---C---I---H--- 332  
Db 23079 KREAAWAFKVTTKCKTLYRISGLVEGTHVFRVLPENLYGIGEBCEPCTSDAVLNVSEVP 23138  
QY 333 --- 335  
Db 23139 LVPKLEVDVDTKSTVTLAWEKPLYDGGSLRTGYVLEACKAGTERWKMVVTLEKPTVLEHT 23198

QY 336 ---L---G---QD-R--P-I---KT--- 344  
Db 23199 VTSLENGOYLFRIRAQNEKGVSEPREFTVAVTQDLRVLPITDLSTMPQKTIHVPAGRP 23258  
QY 345 --- 350  
Db 23259 VELVIPIAGRPPPAASWFFAGSKLRSESRVTVETHTKVAKLTIRETTIRDTGEYTLKLN 23318  
QY 351 --- 358  
Db 23319 VTGTTSETIKVILDKPGPTGPIKIDEIDATISITISWEPPELDGAPLSGVVVEQORDAH 23378  
QY 359 ---P---T-GN---LLAS--C-S--- 368  
Db 23379 RGMPLPVSESVTRSTFKFTRLTEGNEYVFRVAATNRFGIGSYLQSEVIECRSSIRIPGPP 23438  
QY 369 ---D--D-MTL---K-I--W--- 376  
Db 23439 ETLOIFDVSRDGMTLTWYPPEDDGGSQVTGVIIVERKEVRADRWVRVNVKVPVMTTRYSTG 23498  
QY 377 ---S---M---K-Q---DN--- 382  
Db 23499 LTEGLEVEHRTAINARGSKPSPKPIVAMDPIAPGKQONPRVTDTRTSVLSLANSV 23558  
QY 383 ---C---V---H--- 385  
Db 23559 PEDEGSKVTGYLIEWKVDQHEWTKCNTPTPKIREYTLTHLPQGAERYFRVLACNAGGP 23618  
QY 386 ---D---L--- 388  
Db 23619 GEPAEVPVTKVTEMLEYDYELDERYQEGIFVROGGVIRLTIPIKGPFPICKWTKEGQ 23678  
QY 389 ---A---H---NKE---IYTIK--- 398  
Db 23679 DISKRAMIATSETHVELVIKEADRGDSGTYDLVLENCKGKAVY-IKVRIGSPNSPEGP 23737  
QY 399 ---WSP---TG---P---GT--- 406  
Db 23738 LEYDDIQVRSVRSWRPPADGGADILGYLERREVRPKAAWYTIDSRVGTSLVLVGLKE 23797  
QY 407 N---N---P-N--AN---L--- 413  
Db 23798 NVEYHFRVSAENQFISKPLKSEEPVTPKTLNPPPEPSPNPPEVLDVTKSSVLSWSRPK 23857  
QY 414 ---M---L---A--- 416  
Db 23858 DDGSRVTGYIERKETSTDKVVRHNKTIQITTTMYTVTGLVDAEYQPRIIAQNVDVGLSE 23917  
QY 417 -S-AS---F---DS-T---V-R--- 425  
Db 23918 TSPASEPVVCKDPFDKPSQPGLEILSISKDSVTLQWEKPECDGGKEILGYWVEYRQSGD 23977  
QY 426 -LW--- 427  
Db 23978 SAWKSNKERIKDKOPTIGGLEATEYEFVRPAENETGLSRPRRTAMSIKTLTSCEAPG 24037  
QY 428 ---DV---D---R-G--I--- 433  
Db 24038 IRKEMKDVTKLGEAAQLSCQIVGRPLDJKVRFKGKELIQSRKYKMSDGRTHLTVM 24097  
QY 434 ---CI---H---TL--- 438  
Db 24098 EEQDEGVVTCIATNEVGEVETSCKLLQATQFHFPGYPLKEKYKYGAVGSTLRLHVMYIG 24157  
QY 439 ---T---K-H--O--- 442  
Db 24158 RVPAMTWFGQKLLONSENITIENTEHTHLMKVNQVRKTHAGKYKVLQSNVFGTVDAI 24217  
QY 443 ---E---P---V---Y--- 446  
Db 24218 LDVEIQDKPKPTGPIVIEBALLKNSAVISWKPDDGGSWITNYVVEKCEAKEGAWQLV 24277  
QY 447 ---SV---A--F---SP--- 452

Db 24278 SSAISVTTCTIVNLNENAGYFVRVSQNTFGISDPLEVSSVWIKPFEPKPGAPKPTIT 24337  
QY 453 - - - - - DG - - - - - R - - - - - YL - - - - - A - - - - - S - - - - - G - - - - - 460  
Db 24338 AVTKDSVCVAVKPPASDGGAKIRNYILEKKEKKQNKWISVTTBEIRETVFSVKNLEGLE 24397  
QY 461 -SF- - - - - D - - - - - K - - - - - C - - - - - V - - - - - H - - - - - 467  
Db 24398 YEPRVKCENLGGSEWSEISEPITPKSDVPIQAPHPKEELRNLRVYQSNATLVCKVTGH 24457  
QY 468 - - - - - I - - - - - W - - - - - N - - - - - T - - - - - Q - - - - - G - - - - - 474  
Db 24458 PKDIVRWYQGGKEIADGLKYRQEPKGGYHQHIIASVTDGDDATVYQVRATN-QGGSVSG 24516  
QY 475 - - - - - A - - - - - L - - - - - VHSYRG - - - - - T - - - - - G - - - - - 485  
Db 24517 TASLEVEVPAKIHLPKTLGCGAVHALRGFVVSIKIPFSGKPDVITWQGDLDNNGH 24576  
QY 486 - - - - - I - - - - - F - - - - - E - - - - - VC - - - - - 490  
Db 24577 YQVIVTRSFSLVFPNGVERKADAGFYVCAKNRFGIDQKTVELDVADVPDPGRGVKVSDA 24636  
QY 491 - - - - - W - - - - - N - - - - - A - - - - - 493  
Db 24637 SRDSVNLWTEPASDGGSKITNYIVEKCATBRLRVGQARETRYVINLFGKTSYQPR 24696  
QY 494 - - - - - A - - - - - G - - - - - DK - - - - - V - - - - - 498  
Db 24697 VIAENKFGLSKPSEPSEPIITKEDKTRAMNYDEVDRETVSMKASHSTKELYKMYI 24756  
QY 499 - - - - - G - - - - - AS - - - - - A - - - - - SD - - - - - 504  
Db 24757 AEDLGRGEFIVHRCVETSSKTYMAKFKVKGTQDQVLVKKEISILNIAHRNHLHES 24816  
QY 505 - - - - - G - - - - - G - - - - - S - - - - - VC - - - - - V - - - - - LDUR 513  
Db 24817 FESMBELVMIFEFISGLDIPERINTSAFELNREIVSYHVQVCEALQFLHSHNIGHFDIR 24876  
QY 514 - - - - - K 514  
Db 24877 PENIYYQTRRSSTIK 24891

RESULT 7  
T15789  
hypothetical protein C41A3.1 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004  
C:Accession: T15789  
R:Bentley, D.  
submitted to the EMBL Data Library, November 1995  
A:Description: The sequence of C. elegans cosmid C41A3.  
A:Reference number: Z18404  
A:Accession: T15789  
A:Status: preliminary; translated from GB/EMBL/DDBJ  
A:Molecule type: DNA  
A:Residues: 1-7829 <BEN>  
A:Cross-references: UNIPROT:Q18559; EMBL:U41541; NID:g1109867; PID:g1109869; PIDN:AAA831  
C:Genetics:  
A:Gene: CBSP:C41A3.1  
A:Introns: 29/2; 69/3; 141/2; 192/2; 271/3; 321/2; 1000/1; 1098/1; 1127/3; 1168/2; 1282/2; 7514/3; 7596/3; 7654/3; 7716/3; 7732/3; 7769/2; 7797/3  
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein  
P:5547-6616/Domain; acyl carrier protein homology <ACP>  
F:2832,5271,6580/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 70.6%; Score 2590.9; DB 2; Length 7829;  
Best Local Similarity 8.5%; Pred. No. 5e-52;  
Matches 447; Conservative 52; Mismatches 11; Indels 4771; Gaps 356;  
QY 1 MS-----I-----S-----S-D-E-----VN-----9

Db 317 MSPKSMGHGEGAAGLIGVLQSLYSWQHGIIPNQLHLELPSDLGDKSMGFVNEEMELN 376  
QY 10 - - - - - FL - - - - - 11  
Db 377 RVAISSYGGGTNACALIEKPEKPSLVQKESYAESNVLFSLSAKSHESKLQIBEYTOPMA 436  
QY 12 - - - - - V - - - - - Y - - - - - R - - - - - Y - - - - - LOSS - - - - - G 20  
Db 437 QSDSAMEDILYTVNERKTKYDFRAAVFGKDNBEIARKLQGDYSLTLQBSTFEVEFEG 496  
QY 21 - - - - - FSHS - - - - - 24  
Db 497 NEKMLLRMLYKNETF-HSTVDKYCKLAETCFPEPARTALFPFFKLTLLPLTYNVSRLLI 555  
QY 25 - - - - - A - - - - - F - - - - - TF - - - - - G - 29  
Db 556 SSWATFELLVQYNTLKNLGRGKLGQIFCLAVAKVITFESAVQLIKGVAAENLTDILGD 615  
QY 30 - - - - - IE - - - - - S - - - - - HIS - - - - - 35  
Db 616 IELKSSKIPTEIQHLASTKKKILPIHISGELKETAKPNLMTFIVNGBEILELDPVRKVQK 675  
QY 36 - - - - - QSN - - - - - 38  
Db 676 LICQLFACGDPAPKPRGRIVKPTYSFLKQFVPEVQTAMTNIIVDEQTNSSLSDAIE 735  
QY 39 - - - - - IN - - - - - GA - - - - - L - - - - - V - - - - - P - - - - - PAAL-I 50  
Db 736 STVRTIVKQFLDEEDDINLETGAVDSLTSIEMVEAFGTAVNQTFPFDLLEAYPTILNI 795  
QY 51 - - - - - SI - - - - - I - - - - - Q - - - - - 54  
Db 796 VDFLKLTVTPVTKATTSIHKKTSBELSTDINVIACDYQFAGVEGEGELMDTLTSLRT 855  
QY 55 - - - - - K - - - - - GL - - - - - Q - - - - - Y - - - - - V - - - - - EAE - - - - - 63  
Db 856 TGKSIDIRKQCEGDAGLEVGLLKODISMFNDSFFAIAKDEAEFLDPQHRLLNAAVNAL 915  
QY 64 - - - - - VS - - - - - INE - - - - - D - - - - - GT - - - - - 71  
Db 916 EKSGLTSPDADLFLAISAHSEYRALLAEKHINELDERLWGTVHSMVAGRLAVLMGIRGR 975  
QY 72 - - - - - L - - - - - F - - - - - 73  
Db 976 AMIVDTTGVATALEMAVKSIREGKRFATVATSLQISKWLKSLKTLDDHSTNSFSV 1035  
QY 74 DG - - - - - R - - - - - P - - - - - I - - - - - E - - - - - SLSL 83  
Db 1036 DSGGFCRSDGVGVIIILKTAEGDSAVIKISSAKSHHCGAVMTFVSVSSISQLLEAGFSY 1095  
QY 84 I - - - - - D - - - - - A - - - - - V - - - - - M - - - - - 88  
Db 1096 VEGHGTATGAGSAESMAYQKLGSELMSSVRAQFGHCEVASGLIQLMKVSSIKGKIIP 1155  
QY 89 - - - - - P - - - - - D - - - - - VVQT - - - - - R - - - - - Q 97  
Db 1156 SIHVNLTPSEHNRNENIRLPFAVEEKQIDRSAIVSFGITGTKTVTTERVSQLNDNIQ 1215  
QY 98 - - - - - A - - - - - Y - - - - - R - - - - - D 101  
Db 1216 NCYLLPVSARKDGLKKAACLSLIEMIDNSCESLYDISTTLQKQKTNFKWRTAVVGS SHAD 1275  
QY 102 - - - - - KLAQ - - - - - Q - - - - - 106  
Db 1276 VVLKQKFLTSHTNTSLTNWHISTSHSISGCTFFHNIPEFEDHYSMFCHRLRFEPSHNT 1335  
QY 107 - - - - - H - - - - - A - - - - - A - - - - - AA - - - - - A - - - - - 116  
Db 1336 NSIYHLLAVVYALIRVILKHLNTSNFVGVGFNSLIIVAAVDAAPSHVNLNLLHAFANDDV 1395  
QY 117 - - - - - A - - - - - T - - - - - N - - - - - Q - - - - - QGSAKN - - - - - 126  
Db 1396 RWMKRIARDVTISIENKLLNNGEPITTARQAVEATIDQ-KVKNVRLPETTILSPSRA 1454

QY 127 -----GE-----N-----T- 130  
Db 1455 YEPASOLETIQYKUIGEKFTQGGQVDFAGIFGTPIKILDLPEYFNRKSFWLPIPD5V 1514  
QY 131 -AN--GE-----N-----GA-----H----- 139  
Db 1515 PSNEKEKPLIPKSYBFLKSKQWQHVNVD5KIVLPGAT5IRLVHQLNGKPTVELSN 1574  
QY 140 -----T-----I-----A-----N-----N-----HT 146  
Db 1575 IDFLNKITSEAPSVVKIEBQDGLKLVFGETDAISGFKLTELQNFNPINERLNAEVHHT 1634  
QY 147 D-----MM-----E--VD-----G-----DV-----E- 156  
Db 1635 DNIYERFANSHLTYNREFQWDSKYTMKGGEVRFVSMKDLIDLDTLQAIVCYFFEN 1694  
QY 157 -----IP-----S--N-----K--AV-----VL----- 165  
Db 1695 TNDNSPFPTIDQLSILNGDISQQLHAVLVKYDSSGNFINGDATVYDALGNILHISNV 1754  
QY 166 -R--G-----H-----ESE-----V----- 172  
Db 1755 TFXKLGQASPLSTKTVDS5KITKVENEDOKRASKMHLHWFE-ENFGWTDIDNTGPF 1813  
QY 173 F-----I-----C-----W----- 175  
Db 1814 FDLGLTSIQAVKLUNA5KSNYPNASSTCFDYP5IDL5G5LTLNDPQVTTSTGEDDI 1873  
QY 176 -----A-----W----- 177  
Db 1874 QKDLTEDHKTRLAENPIGWMAAACRLPGCVSPSELWELKIGKKNASSRIPATRVPTRN 1933  
QY 178 -----NPV-----S-----D-----LL--V-----SG- 187  
Db 1934 TLISGKYGNFVGGNFIQDVTQDP5PPFKISKEAELIDPQORLLECVQCLENSGV 1993  
QY 188 -S-----G-----D-----ST-----A-R--I--W----- 196  
Db 1994 IETSNVGFVGLMEKEYQDM5ESS5LAMLG5MAAVIAGRVNIFGCGPSVTIDTACSS 2053  
QY 197 -----NL----- 198  
Db 2054 SLVALEMAINALLDNCR5KVIIVAGVNLILNEK5QGLRTNGKML5QHGM5LSP5RASGYG 2113  
QY 199 -S-----E--N-----ST--S-----G-----P-----T-----Q----- 208  
Db 2114 RSDGCVVLMELAKPNFHYMSTIQSVNVN5HGR5SVSLTAPNGVAHKMLLT5VINQ5PSLA 2173  
QY 209 -----L--VLR-----H-----C----- 214  
Db 2174 IDYEAHGCTPLGDPFIEFTL5SILQNIIGSVKASL5GHEASAGTCGLLKLFLMLTYQ 2233  
QY 215 -----IR-----E-----B-----G--G----- 219  
Db 2234 YVPTLHFVNLKDNAG5IRLPIIGED5ELV5AGIS5FVG5GTNA5A5AFNDNNKLEPY 2293  
QY 220 -----Q-----DVP-----S 224  
Db 2294 IPIHKYILPISAKNQISL5DNLEK5L5VPLTDVDPICN5ASALANNRSHFTIRNALIV5 2353  
QY 225 N-----KD-----V----- 228  
Db 2354 NSGIVN5KMEGKPHRVAKDORYHVKLC5L5D5L5L5QYOVINETYTVASLKNP5Q5PAMKF 2413  
QY 229 -----TSL-----D-----W--N--S--E--GTL-----A----- 241  
Db 2414 AIKFLTSLSEYIEIVASDGEELLAVLLANGSLKWNFNKMTIELPIG5LTLTEFADHDLN 2473  
QY 242 -----TG-----S-----Y----- 245  
Db 2474 STTSS5IK5YQTQ5P5SHN5L5D5PMELMKLIMKLYITGYDV5DWATVY5PVEQFIALPNYQFN 2533

QY 246 -----D--G-----F----- 248  
Db 2534 KQTLWFEERLEIVD5HVLIGTIDE5EDTILLLKNQISELRHPQFFKPKPLDVGTMSEIAIE 2593  
QY 249 A--R-----I-----W-----T----- 253  
Db 2594 ALKIRNEIPFSIQNLKTELITLTKPAWLETNTVTRNEDDEGNVSAVIDGQRLFSLNASSV 2653  
QY 254 -----K----- 254  
Db 2654 EIQNI5EVP5VEQIPDKVVYLKECPNAVIRRHNMVYVDSRABQSPFRTANIVLNEIIGF 2713  
QY 255 -----DG-----N----- 257  
Db 2714 APTPSDMFIEILGVLPSVHYMVQVDDGALWQFQMISQDKRVL5NIYVLKDAKLEIPTIR 2773  
QY 258 -----LA-----ST-----LG----- 263  
Db 2774 MHK5STLLSSQ5EASIVAAKTLQMAVRHKVCLAVGDVIESGLDID5Q5L5TG5F5ELGID5L 2833  
QY 264 ----- 263  
Db 2834 ATVDLLNRLNQKYFPEIELT5DLFDNPSIIDLSIMIEQLLNEK5GITEP5EPNTPKTSLR 2893  
QY 264 -----Q-----H----- 265  
Db 2894 GRKLSIPAVRAQVLAQIEFVENYN5KQ5EQA5PSS5SEC5NHL5E5DATVDRTEIRRK 2953  
QY 266 -----KG-----P-----IF----- 270  
Db 2954 VSLAVFDLATETLSA5DLQ5K5GTELGMDSLSIVDFVNR5LNDKYFPDDEITASDIFDPT 3013  
QY 271 -----A----- 271  
Db 3014 VDELS5DHIVR5K5SSVPPA5A5IMK5TMNGIST5VDA5HTK5LEN5QS5FML5N5Q5INP 3073  
QY 272 -LK--W--N--K--K--GNFI--LSA-----G-----VD--K-----T-T- 290  
Db 3074 TLKMI5NQTIKLVKPSDGNFLFELNANGQ5E5K5QKHFTGPN5NIIDLKGFHEG5TETL 3133  
QY 291 -----I-----I--W----- 293  
Db 3134 YMSLLN5K5SISKLEIQCRF5G5Q5E5F5GL5G5NISRA5F5MKT5VA5E5KN5PLIS5P5W5Q5V5 3193  
QY 294 -----D-----A-----H 296  
Db 3194 FVDS5D5PITGNW5LITG5L5G5IG5LEIG5K5FIANNGA5EN5LIS5RQ5PTAKALR5DIL5TELTH 3253  
QY 297 TG----- 298  
Db 3254 IGLAKTIVL5K5IKNISAKLIVFQ5P5KL5SF5TK5KYISDL5F5SK5V5TF5YN5LLQ5K5F5KH5F 3313  
QY 299 -----E-----A-----K--Q-- 302  
Db 3314 HMLF5EW5LV5LIQ5W5VHT5IAD5IND5KE5L5IRE5L5TK5LNV5GIT5G5I5H5AG5VL5K5DK5K5IERQ5N 3373  
QY 303 -----Q-P-P--FH----- 307  
Db 3374 KESFNQVFTPKANG5FHV5LEE5K5HFNYK5IEN5FIM5SS5FTAAC5G5N5Q5L5N5GV5N5AY5LEYQ 3433  
QY 308 -----S-----A-----PA 311  
Db 3434 VQRR5RQ5G5G5CAIQ5NG5WID5TGM5D5M5D5N5VR5K5FLAN5GLF5G5H5K5D5AL5K5YL5R5CIL5TK5PE 3493  
QY 312 L-----DVDW-----Q-----SN----- 319  
Db 3494 LIWVANID5W5IL5KN5K5DL5PK5LINT5GIL5P5FD5TG5M5E5P5PL5SG5D5EK5V5M5N5F5VE 3553  
QY 320 -----N-TF----- 322  
Db 3554 DEBE5VLE5LIK5E5V5S5IL5M5CS5PT5KL5KN5KN5IM5DM5GL5K5LIVE5FLN5FIN5T5FK5IS5VN5L5DA 3613  
QY 323 -----A-----S--C-----ST 327

Db 3614 YNPTLEKLAHIFEQMTIVDHPVNSVKSEBEIFKSTDFCFIFGINIFFDKNDKDFRAKST 3673  
QY 328 -----D----- 328  
Db 3674 AVKLENGEOLPTAGKYAVSVGKSRDVVSKIKKAPQOIKLCQSSSKCVLMLTGQGSQ 3733  
QY 329 -----M-----C-----H--V-- 333  
Db 3734 YPMGRQLVNEYEIFRTTLQCLCKCDEYLQGVSLWEILLFNTDHYKLLQLTKHMQPIMF 3793  
QY 334 C-----K-----LG 337  
Db 3794 CFGYATAQLWLSLGIYPDYLLGHVSGELVAGVLGIMSIEDGLRLIVERGKAMENIAGLG 3853  
QY 338 -----O--D--R-----P--I--KT-----F--QG----- 347  
Db 3854 ALLAVOREIADEVLRKPKVSVATINSGKVVFAGTKSVLDAALAFVKGQKQATYVNOQY 3913  
QY 348 --HTN--E-----VN--A--I----- 355  
Db 3914 PFHSNLIQETHLVSLRQCLADIKFSAGRTPLVSNVTQIINTFSEAVIVKHTVSAVKFVD 3973  
QY 356 -----K--W--D-----PT-----GNLL--A-- 365  
Db 3974 CVETLOAKGVTVWIDAGSAVLATFVKRIITOPTELSKHRIVQTCKEKESVDVNLVQACLE 4033  
QY 366 -----S--C-----S----- 368  
Db 4034 LEOSGLPISTWTLTYGCGRNADERLVEFFPMTHNDIINKDFELLEGHQLNGKIVVAGAYQL 4093  
QY 369 -----D-----DMT----- 372  
Db 4094 FKIDQLVKLKAAGMELMKNVFLKWPYIEDNREYQIQWNSDMTIELIVNSVIVCSLEVE 4153  
QY 373 --LK-----I----- 375  
Db 4154 PQNSVLKLETISENEKPFVHDFYETLFRNLQYDGFRRIESARRDKRCFQIKSSPF 4213  
QY 376 -W-----S--MK-----Q-----D-- 381  
Db 4214 AMPLIDSAMHSITASVVRPRPCYFLPVANGSVTMKDTNSTFLPNLHAQTVITSETDKFI 4273  
QY 382 --N-----C-----VH-----D--L-- 387  
Db 4274 QVNVALLAGDTPICEVNMNTIIVLKTPEVHTRIPIINSIEVETPPKSEIEIVGDISLPYN 4333  
QY 388 Q-----AH-----N--K--I-----Y----- 395  
Db 4334 QISENSENHQKLTNTVKQKLHNSLQDHARVALLDSADARYWDPEYFGIRPSEAKFIDP 4393  
QY 396 -----T-----I-----K--WS----- 400  
Db 4394 QORLLCSVAKLDSLILTSNTGTGVFGCSANEFSHVIYAYGYKDPRAEWSGGTSNSA 4453  
QY 401 -----P-----TG----- 403  
Db 4454 LAGRIAHWLKLGPPVTLDTACSSSFYALSACDALRTGCEYAIIVGTNVNLVHMTVDV 4513  
QY 404 -----P----- 406  
Db 4514 LQNAKMTVDDFCAPVDANGYKRSBVCMSMLTKSPNIDSVATITNYATGHNGTSSSLF 4573  
QY 407 --N-----NP-----NA-----N-- 412  
Db 4574 TPNGLSOLEVMORATNPLEKILBIOTHCTGTGLGDPRIEINATSKLVSSACKIGSVKSNIG 4633  
QY 413 -----LM----- 414  
Db 4634 HTEGSSGLVSLCSSLSFRRSKRYVAQHLKCPNTSNIKTNMICRPFGEADENNLSILNN 4693  
QY 415 -----L--A-----SA-S-----F----- 420

Db 4694 FGFTGSCSVLKPKNNAISEHFVSSEVFPYILLSSHSAKSQKYQVQVLCFISNSAKSLH 4753  
QY 421 D--S-----T-----VR----- 425  
Db 4754 DIMMSLFQKKIHVHROPIIFNFKRKIAVTSLDGFVEVRDERLSKLKHPCSVFLKEGVVHF 4813  
QY 426 -----L-W-----D-VDR-----GI----- 433  
Db 4814 DKDKSFQVDLPISIVFNNTLHWALDSYRDEIDRHESQOMSFKNIFYEKVILETMPQONP 4873  
QY 434 --CI-----H-----TL----- 438  
Db 4874 ISKVCICIRGLDLPKIEIDEVSSFANGIIIVFHPISNSIFEYLKAKVWSLISRNQNVFI 4933  
QY 439 -----TKH-----Q--E--P-----V-----YS----- 447  
Db 4934 ICCFENGTSHTWTGTLRSLSASEKMIYPKPVSDIKVDALKLEPNHEMFMFAIYFKRGY 4993  
QY 448 -----V--AF-----S-----P-----DG----- 454  
Db 4994 VERLRKVNPTFKPAQOYKELISGGTGGIGSAIINELKPKSSVITITKNTIASEDGKTFLS 5053  
QY 455 --R-----Y--LASG-----SPD-----KC----- 465  
Db 5054 SDITRLDISHKFNYPVFLA-GIVNNSLHENVKRDSLDEMSIKLOGAKNLMKCCDETSHF 5112  
QY 466 -----VH----- 467  
Db 5113 VFSSSIANVLGSGQSNYAFNSGLVTSFLETSTKSTIIHWGPWKDVGMLAQPERREIVK 5172  
QY 468 -I--W--N--TQ--T-----G--ALV--H----- 478  
Db 5173 QIESNGMKLLPNQDAISVFYTFQMETHEQIIVFDGDFDITVARQPHLQKLLSEVVEKTKV 5232  
QY 479 -----S----- 479  
Db 5233 KBEIKKSLNFBIEIFEIVGITDISKLNIPFMDLIGIDSLCMENLRYSLNKNFDELTVS 5292  
QY 480 -----Y-----R----- 481  
Db 5293 EMPENATYKLOQYVETLRKAKHSLSHDRVSSQVSNKEDDTRVAVIGWSABFSGSSNI 5352  
QY 482 -----G--TG----- 484  
Db 5353 HEYWNELMDGICSTGNKYLLKNPFGDNKFNLTDEDARVLDPOVRKFIQHAYLALENS 5412  
QY 485 -----GIF--E-----V-----C----- 490  
Db 5413 GYVKQKHELRCGVFAGABPSDYGRADDHDDAMRKLFVMNMNSYLSASYCLDLKGEAVS 5472  
QY 491 -W-----NA-----AG--D-----KV-GA-----SAS----- 503  
Db 5473 VYSACSTALVANAVKSIQSGMDYALVGAASIAEVSALSGDFDQKTMFKSGVCRP 5532  
QY 504 --D-----GS-V-C-VL-----DL-----R----- 513  
Db 5533 FDKDSBEGIVRSGVGVFLKRYSQLLDNDNVHFVVKDFAINNDGHSRASFMAPNAGQL 5592  
QY 514 K 514  
Db 5593 K 5593

## RESULT 8

S25111

alpha-2-macroglobulin receptor precursor - mouse

N:Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor

C:Species: Mus musculus (house mouse)

C:Date: 04-Sep-1998 #sequence\_revision 04-Sep-1998 #text\_change 09-Jul-2004

C:Accession: S25111; S32554

R:Van Leuven, F.

submitted to the EMBL Data Library, July 1992

A:Reference number: S25111

A;Accession: S25111  
A;Molecule type: mRNA  
A;Residues: 1-4545 <VANI>  
A;Cross-references: UNIPROT:Q61291; EMBL:X67469; NID:g49941; PIDN:CAA47817.1; PID:g49942  
R;van Leuven, F.; Stas, L.; Raymakers, L.; Overbergh, L.; de Strooper, B.; Hilliker, C.;  
Biochim. Biophys. Acta 1173, 71-74, 1993  
A;Title: Molecular cloning and sequencing of the murine alpha-2-macroglobulin receptor c  
A;Reference number: S32554; MUID:93250049; PMID:8485155  
A;Accession: S32554  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-28,4416-4453 <VAN2>  
A;Cross-references: EMBL:X67469  
C;Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated  
ciated protein (see PIR:JX0281).  
C;Superfamily: Alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding  
C;Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; glycoprotein  
F;1-19/Domain: signal sequence #status predicted <WAT>  
F;20-3944/Domain: alpha-2-macroglobulin receptor 515K chain #status predicted <S15K>  
F;28-65/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F;73-109/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F;116-149/Domain: EGF homology <EG1>  
F;155-189/Domain: EGF homology <EG2>  
F;199-240/Domain: LDL receptor WYTD-containing repeat homology <YW01>  
F;241-282/Domain: LDL receptor WYTD-containing repeat homology <YW02>  
F;293-335/Domain: LDL receptor WYTD-containing repeat homology <YW03>  
F;336-379/Domain: LDL receptor WYTD-containing repeat homology <YW04>  
F;380-421/Domain: LDL receptor WYTD-containing repeat homology <YW05>  
F;422-469/Domain: LDL receptor WYTD-containing repeat homology <YW06>  
F;479-520/Domain: EGF homology <EG3>  
F;572-614/Domain: LDL receptor WYTD-containing repeat homology <YW07>  
F;615-660/Domain: LDL receptor WYTD-containing repeat homology <YW08>  
F;661-711/Domain: LDL receptor WYTD-containing repeat homology <YW09>  
F;712-753/Domain: LDL receptor WYTD-containing repeat homology <YW10>  
F;754-800/Domain: LDL receptor WYTD-containing repeat homology <YW11>  
F;808-843/Domain: EGF homology <EG4>  
F;855-891/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
F;896-932/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
F;937-972/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
F;977-1012/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
F;1016-1052/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
F;1063-1098/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F;1105-1141/Domain: LDL receptor ligand-binding repeat homology <LDL9>  
F;1146-1183/Domain: LDL receptor ligand-binding repeat homology <LDLA>  
F;1186-1222/Domain: EGF homology <EG5>  
F;1228-1262/Domain: EGF homology <EG6>  
F;1270-1309/Domain: LDL receptor WYTD-containing repeat homology <YW12>  
F;1310-1356/Domain: LDL receptor WYTD-containing repeat homology <YW13>  
F;1357-1399/Domain: LDL receptor WYTD-containing repeat homology <YW14>  
F;1400-1446/Domain: LDL receptor WYTD-containing repeat homology <YW15>  
F;1447-1489/Domain: LDL receptor WYTD-containing repeat homology <YW16>  
F;1490-1532/Domain: LDL receptor WYTD-containing repeat homology <YW17>  
F;1541-1579/Domain: EGF homology <EG7>  
F;1584-1627/Domain: LDL receptor WYTD-containing repeat homology <YW18>  
F;1628-1670/Domain: LDL receptor WYTD-containing repeat homology <YW19>  
F;1671-1714/Domain: LDL receptor WYTD-containing repeat homology <YW20>  
F;1715-1757/Domain: LDL receptor WYTD-containing repeat homology <YW21>  
F;1755-1797/Domain: LDL receptor WYTD-containing repeat homology <YW22>  
F;1798-1847/Domain: LDL receptor WYTD-containing repeat homology <YW23>  
F;1851-1887/Domain: EGF homology <EG8>  
F;1935-1977/Domain: LDL receptor WYTD-containing repeat homology <YW24>  
F;1978-2020/Domain: LDL receptor WYTD-containing repeat homology <YW25>  
F;2021-2064/Domain: LDL receptor WYTD-containing repeat homology <YW26>  
F;2065-2106/Domain: LDL receptor WYTD-containing repeat homology <YW27>  
F;2107-2152/Domain: LDL receptor WYTD-containing repeat homology <YW28>  
F;2160-2195/Domain: EGF homology <EG9>  
F;2200-2242/Domain: LDL receptor WYTD-containing repeat homology <YW29>  
F;2254-2295/Domain: LDL receptor WYTD-containing repeat homology <YW30>  
F;2345-2389/Domain: LDL receptor WYTD-containing repeat homology <YW31>  
F;2390-2430/Domain: LDL receptor WYTD-containing repeat homology <YW32>  
F;2431-2474/Domain: LDL receptor WYTD-containing repeat homology <YW33>  
F;2483-2518/Domain: EGF homology <EG10>

F;2525-2562/Domain: LDL receptor ligand-binding repeat homology <LDLB>  
F;2567-2601/Domain: LDL receptor ligand-binding repeat homology <LDLC>  
F;2606-2640/Domain: LDL receptor ligand-binding repeat homology <LDLD>  
F;2653-2689/Domain: LDL receptor ligand-binding repeat homology <LDLE>  
F;2697-2731/Domain: LDL receptor ligand-binding repeat homology <LDLF>  
F;2735-2770/Domain: LDL receptor ligand-binding repeat homology <LDLG>  
F;2775-2813/Domain: LDL receptor ligand-binding repeat homology <LDLH>  
F;2819-2854/Domain: LDL receptor ligand-binding repeat homology <LDLI>  
F;2859-2898/Domain: LDL receptor ligand-binding repeat homology <LDLJ>  
F;2905-2940/Domain: LDL receptor ligand-binding repeat homology <LDLK>  
F;2945-2981/Domain: EGF homology <EG11>  
F;2987-3022/Domain: EGF homology <EG12>  
F;3030-3069/Domain: LDL receptor WYTD-containing repeat homology <YW34>  
F;3070-3114/Domain: LDL receptor WYTD-containing repeat homology <YW35>  
F;3118-3157/Domain: LDL receptor WYTD-containing repeat homology <YW36>  
F;3158-3201/Domain: LDL receptor WYTD-containing repeat homology <YW37>  
F;3202-3242/Domain: LDL receptor WYTD-containing repeat homology <YW38>  
F;3243-3285/Domain: LDL receptor WYTD-containing repeat homology <YW39>  
F;3295-3331/Domain: EGF homology <EG13>  
F;3335-3370/Domain: LDL receptor ligand-binding repeat homology <LDLL>  
F;3375-3409/Domain: LDL receptor ligand-binding repeat homology <LDLM>  
F;3414-3449/Domain: LDL receptor ligand-binding repeat homology <LDLN>  
F;3454-3490/Domain: LDL receptor ligand-binding repeat homology <LDLO>  
F;3495-3532/Domain: LDL receptor ligand-binding repeat homology <LDLP>  
F;3537-3571/Domain: LDL receptor ligand-binding repeat homology <LDLQ>  
F;3576-3610/Domain: LDL receptor ligand-binding repeat homology <LDLR>  
F;3614-3648/Domain: LDL receptor ligand-binding repeat homology <LDLS>  
F;3655-3691/Domain: LDL receptor ligand-binding repeat homology <LDLT>  
F;3696-3732/Domain: LDL receptor ligand-binding repeat homology <LDLU>  
F;3742-3777/Domain: LDL receptor ligand-binding repeat homology <LDLV>  
F;3786-3823/Domain: EGF homology <EG14>  
F;3829-3861/Domain: LDL receptor WYTD-containing repeat homology <YW40>  
F;3869-3912/Domain: LDL receptor WYTD-containing repeat homology <YW41>  
F;3913-3970/Domain: LDL receptor WYTD-containing repeat homology <YW42>  
F;3945-4421/Domain: 85K chain extracellular #status predicted <EXT>  
F;3971-4013/Domain: LDL receptor WYTD-containing repeat homology <YW43>  
F;4014-4057/Domain: LDL receptor WYTD-containing repeat homology <YW44>  
F;4058-4100/Domain: LDL receptor WYTD-containing repeat homology <YW45>  
F;4101-4143/Domain: LDL receptor WYTD-containing repeat homology <EG16>  
F;4152-4183/Domain: EGF homology <EG17>  
F;4201-4232/Domain: EGF homology <EG18>  
F;4237-4268/Domain: EGF homology <EG19>  
F;4273-4304/Domain: EGF homology <EG20>  
F;4309-4340/Domain: EGF homology <EG21>  
F;4345-4375/Domain: EGF homology <EG22>  
F;4378-4409/Domain: EGF homology <EG23>  
F;4422-4445/Domain: transmembrane #status predicted <TMM>  
F;4446-4545/Domain: intracellular #status predicted <INT>  
F;167,2999/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F;2959/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted  
F;4076,4136,4279/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 70.5%; Score 2587.7; DB 1; Length 4545;  
Best Local Similarity 10.0%; Pred. No. 2e-52;  
Matches 431; Conservative 59; Mismatches 13; Indels 3813; Gaps 386;  
QY 1 MS-----I-----SSDE-----V-----N--F-L-----11-  
Db 89 MSRLNGIQCMGDSGEGHCRRLRANCRMGQHCVCPTPTSGTCVNSFFLEADGKT 148  
QY 12 -----VY-----R-----Y-LQ-----E-----18  
Db 149 CKDFECVSYGTQCLCTNTDGSFTGCGVEGYLLOPNRSCRKARNEFVDRPPVLLIANQ 208  
QY 19 -----SG-----FS-----H--SAF-T-----F 28  
Db 209 NILATYLSGAQVSTITPTSTQTTAMDFFSYANETVCVWVHVGDSAAQTOLKARMPGLKGF 268  
QY 29 GI-ESH---IS-----Q-----S-N-----I-----NG-----41  
Db 269 -VDE-HTINISLSLHHVEQMAIDWLITGNFFVDDIDRIFVPCNRNGDTCVTLLDLELYNP 326

Qy 42 --AL--VP-----P-----A- 47  
Db 327 KGIALDPAMKGVFTDYGQIPKVERCDMGQNRKTLVDSKIVPHGITLDSRLVYTWAD 386  
Qy 48 A--LI--SIIQK-L-Q--Y--V-E-----A-----EVS-I----- 66  
Db 387 AYLDIEVDYEGKGRITIIQ-GILIEHLVGLTVFENLYATNSDNANTOQKTSVIRVNR 445  
Qy 67 -N-E--D--GTL----- 72  
Db 446 FNSTEYQVTVRDKGALHIYHQRPVRSHACENDQYKPGGCSICILLANSHKARTC 505  
Qy 73 --F--DG-----R-----PIE-----SL 81  
Db 506 RCRSGFSLGSDGSKCKPEHELFLVYKGRPGIIRGMGMGAKVPDEHMIPIENLMNPRAL 565  
Qy 82 -----S-LI-----D-----AV-M-----PD 90  
Db 566 DPHAETGFIYFADTTSYLIQRKIDGTERETILKGIHNVGVAVDMGMDNLYWTDGPK 625  
Qy 91 --V-V--O--O-----Y----- 99  
Db 626 KTISVARLEKAAQTRKTLIEGKTHPRAIVVDPPLNGHMYTWDWEDPKDSRGRGLERAWM 685  
Qy 100 --RD--K-----L--A--Q-----Q-----H 107  
Db 686 DSHRDIFVTSKTVLWPNGLSLDIPAGRLYWVDAFYDRIETILLNGTDRKIVYEGPELNH 745  
Qy 108 A-----AAA-----A----- 112  
Db 746 AFGLCHHGNYLFWTEYRSGSVYRLERGVAGAPPTVTLRSERPPIPEIRMYDAHEQOQVT 805  
Qy 113 -----AA-----A-A-----AT--N-----Q----- 120  
Db 806 NKCRVNGGSSICLATPGSRQACAEQVLDTDGTCLANPSVPPQCPQCFACANN 865  
Qy 121 --Q-----G-----S--A-----K-----N--G----- 127  
Db 866 RCIOERWKCDGNDCLDNDSEAPALCHOHTCPSPDRPKCENNRCPINRWLCGDGNDGENSE 925  
Qy 128 -E-N-T--A-----N-----G-----E----- 135  
Db 926 DESNATCSARTCPNPFQSCASGRCPISWTCDLDDCGDRSDESASCAYPCTPPLTQFTC 985  
Qy 136 -NG-----A--H--T-----I----- 141  
Db 986 NNGRCININWRCDNDGNSDEAGCSHSCSTQFKNSGRCPPEHWTCDGNDGCDYS 1045  
Qy 142 --AN--N-----HTDME-----V-----DGDV-----E----- 156  
Db 1046 DETHANCTNQTAPPGGCHSD--EFQCPLDGLCIPLRWRCGDTDCMDSSEKSCGVT 1103  
Qy 157 I--PSN--K--A--V-----V-----L--R----- 166  
Db 1104 VCDP--NVKFGCKDSARCISKAIVCDGSDCEDNSDENECALACRPPHPCANNTSVCLP 1162  
Qy 167 -----G-----H-----E-----S-----E----- 171  
Db 1163 PDKLCKDKDCGDSDEGELCDQCSLNNGCSHNCVAPGEGIVCSPLGMEIGSDNHTC 1222  
Qy 172 -V-----F-I-CA--W--NP--V--S--D--LLV--S----- 186  
Db 1223 QIQSYCAKHLKCSQKQDONKFSVKCSYEGWVLEPGETCRSLDPFKLFIIFSNRHEIR 1282  
Qy 187 -----G--S-----D-----S-----T-A-----R-----I-- 195  
Db 1283 IDLHKGDYSVLVPEGLANTIALDFHLSQSALYWTDAVEDKIIYRGKLLDNGALTSFEVVIQY 1342  
Qy 196 -----W--N--L--SE--N-----ST--SG-----PTQ--L-----V 210  
Db 1343 GLATPEGLAVDWIAGNIYWVESNLQIEVAKLDGTLRLLAGDI BHP-RAIALDPRDGI 1401  
Qy 211 L-----R--H-----C-----I-- 215

Db 1402 LFWTMDASLPRIEAAASMGAGRRTIHRETSGCGCANGLTVDYLEKRIILWIDARSDAIS 1461  
Qy 216 -----R--E-----GG----- 219  
Db 1462 ARYDGSCHMEVLRGHEFLSHPPFAVTLYGGBEVYWDRTNTLAKANKWTHGNVTVVQRTNT 1521  
Qy 220 Q--D--V--PS-----N-----K----- 227  
Db 1522 QPFDLQVYTHPSROPMAPNPCEANGRGPCSHLCLINYNRTVSWACPHMLKHLKNDTTCYE 1581  
Qy 228 -----N-SEG--T-----L--LA-----T-----GS----- 244  
Db 1642 RAFINGT-GVETVVSADLPNAHGLAVDWVRNLFWTSYDTNKKQINVARLDGSPNAVQ 1700  
Qy 245 -----Y-----DG-----F-----A----- 249  
Db 1701 GLEQPHGLVHPLRGKLYWTDGDNISMANMDGSNHTLLFSGQKGPVCLADFPESKLYWI 1760  
Qy 250 -----R-----I-----W-----T--K--DG-- 256  
Db 1761 SSGNHTINRCNLGSELEVIDTMRSQLGATALAIMGDKLWADQVSEKMGTCNKADGSG 1820  
Qy 257 -----N--LA-----I-----S-----T-----L 262  
Db 1821 SVVLRNSTTLVMMHMYVDESIOLEHEGTNPCSVNNGDCSOLCLPTSETTSCMCTAGYSL 1880  
Qy 263 --CQ-----HKG-----P-----I--F--A----- 271  
Db 1881 RSCQACEGVGSPLLSVHEGIRGIPDPNDKSDALVPVSGTSLAVGIDPHAENDTIYVW 1940  
Qy 272 --L-----K--W-----N-----K-----KGN-----F--I--L----- 282  
Db 1941 DMGLSTISRAKRDQWREDVVTNGIGRVEGIAVDWIAIGNIYWTDOQGSFVIEVARLNGSFR 2000  
Qy 283 --SAGVDK-----T-----T-----I-----I-- 292  
Db 2001 YVVISQGLDKPRAITVHPEKGYLFWTEWGHYPRIERSRLDGTERRVVLNVNVSISWPNGISV 2060  
Qy 293 -----W-DA-----H----- 296  
Db 2061 DYGGKLYWCDAEMDKIERIDLETGENREVLSNNMDPSVSPFEDFIWSDRTHANGS 2120  
Qy 297 -----TG----- 298  
Db 2121 IKRGCKDNATDSVPLRTGTIGVQLKIKVFNDRQKGTNVCANGGQQCLLYRGGQRA 2180  
Qy 299 -----E-----A-----K----- 301  
Db 2181 CACAHGMLAEDGASCREYAGYLLYSERTILKSIHLSDERNLNAPVOPFEDPEHMKNVIAL 2240  
Qy 302 -----Q----- 303  
Db 2241 AFDYRACTSGPTNRIFFSDIHFNIQIINDCGSRTTIVENVGSVEGLAYHRGWDTLYW 2300  
Qy 304 -----F-----P--F-----H----- 307  
Db 2301 TSYTTSTITRHTVDTQTRPGAFERETVITWGGDDHPRAFVLDECONLMFWTNWNLHPSIM 2360  
Qy 308 --S--A-----P-----A-LD----- 313  
Db 2361 RAALSGANVLTIEKDIRTPNGLAIDHRAEKLYFSDATLDKIERCEYDGSHRVILKSEP 2420  
Qy 314 -----V-----DW-----Q----- 317  
Db 2421 VHPFGLAVGEHIFWTDWVRRAVORANKYVGSMDKLLRVDIPOPPIIIVANDTNSCEL 2480  
Qy 318 S-----NN-----T-----F--A-----SC-STD--M--CI-- 331

Db 2481 SPCRINNGCQDLLTHQGHVNCRCGRILQEDFTCRVAVSSCRAQDEFECANGECIS 2540  
QY 332 ---H---V--- 333  
Db 2541 FSUTCGVSHCKDKSEKPSYCNRRCKTFRQNNRCVSNMLWNCVNDYCGDGSDEIP 2600  
QY 334 C-K---LG---Q- 338  
Db 2601 CNKTACGVGEPRCDSCIGNSSRCNQFVDCEDASDEMNCSATDCSSYFRLGVKGVLP 2660  
QY 339 ---D---RP---I---K- 343  
Db 2661 CERTSLYAPSWYCDGANDCGDYSDERDCPVKRCPLNYFACPSGRCPMSWTCDKED 2720  
QY 344 ---T---F---Q--- 346  
Db 2721 DCENGEDETHCNKFCSEAFECQNHRCISQWLCDGSDDCGDSDEAAHCEGKTCGPSSF 2780  
QY 347 ---G-H---T---N---E---V--- 352  
Db 2781 SCFGTHVCVPERWLCDGDKDCTGDADESVTAGLYNSTCDDREFMCQNRLCIPKHVCDH 2840  
QY 353 ---N---A---I---KW---D---P-T- 360  
Db 2841 DRDCADGSDSPCEVPTCGNEFRCAANGRCCLSRQWECDEGNDCHDHSDEAPKNPCHTS 2900  
QY 361 ---D---M---T-LK---I-W--- 376  
Db 3021 SCAVTEDEEPLIFANRYLRKLNLDGNSVTLKQGLNNAVALAFDYREQMIYWTGTTQ 3080  
QY 377 -SM--K---Q---D---N---C--- 383  
Db 3081 GSMIRHNLGNSVQVHLRTGLSNPDGLAVDWGGLYKDCDKGRDTEVSKLNGAVRTVL 3140  
QY 384 ---V---H--- 385  
Db 3141 VSSGLREPRALVDVQNGYLYWTDWGDHSLIGRIGMDGGRSIIVDTKITWPNGLTVDYV 3200  
QY 386 ---D---L-Q--- 390  
Db 3201 TERIYADAREDYIEFASLDGNSRHHVLSQDIPIHIFALTFLFYVYVWTDWETKSINRAHK 3260  
QY 391 ---NK---EI--- 394  
Db 3261 TTGANKTLTSLHRPMDLHVHALRQDPVNPHPKYNNGGCSNLCILSPGGHKACAPT 3320  
QY 395 -Y---T---IK-W---S---P--- 401  
Db 3321 NFYLGGDRTCVSNCTASQFVKNDKCIPPWKKDTEDDCGDHSDEPPDCPEKRCRQGF 3380  
QY 402 ---TG---P---G---TN---N- 408  
Db 3381 QCSTGICTNPAFICDGDNDQDNDSEANCIDHVLCPSPQFKTNTNRICPGIFRCNGQDNC 3440  
QY 409 ---PN---ANL-M---L 415  
Db 3441 GDGEDERDCPEVTCAPNQFQCSITKRCIPRWWCDRDNHCVDGSDPEANCTQTCGVDEF 3500  
QY 416 ---AS---A---S---F--- 420  
Db 3501 RKCDSGRCIPARWKCDGDDCGDSDPEKBECDERTCEPYQFRCKNNRCVGRWQCYDN 3560  
QY 421 ---D-S-T---V---R--- 425  
Db 3561 DCGDNDSEESCTPRPCSESEBFFCANGRCIAGRWKCDGHDCAHGSDKDCDTPRCMDQFQ 3620

QY 426 ---L-W---DVD---R--- 431  
Db 3621 CKSGHICILPFWCDDADACMDGSDBEACGTGVRTCPLDFEQNNLTCKPLAWKCDGEDDC 3680  
QY 432 G---IC---I--- 435  
Db 3681 GDNSDENPECARFICPPNRPFRCKNDRVCLMTGRQCDGVNCGDGTDEBCEPPTAQN 3740  
QY 436 H---T---L--- 438  
Db 3741 HCKDKKEFLCRNQCLSSSLRCNMFDDCGDSDDEDCSIDPKLTSATNASMCGDEARCV 3800  
QY 439 -T-K---H---Q-E---P--- 444  
Db 3801 RTEKAAACACRSGFHTVPOPGQDINECLRFGTCSQLWNKPKGGHLCSCARNFMKTHNT 3860  
QY 445 ---V-YsVA---FSP---D---GR--- 455  
Db 3861 CRAEGSEYQVLY-IADDNEIRSLF-PGHPHSAYEQTFOGDESVRIDAMDVHVKAGRVYWT 3918  
QY 456 ---Y-L---A---S---G---S---F--- 462  
Db 3919 NMHTGTISYRSLPAPPTTSNRHRQIDRGVTHLNLISGLKMPRGIAIDWVAGNVYWTDS 3978  
QY 463 -D---K-C---V---H-I---W-N--- 470  
Db 3979 GRDVEIAQMKGNKRLTISGMIDEPHAIIVDPLRGTMYSWDGNHPKIEAAMDGLRE 4038  
QY 471 T---Q-TGAL-V-HS---Y---R-GT---G--- 484  
Db 4039 TLVQDNIQWPTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPIVAADSKRGLSHPPSID 4097  
QY 485 ---G---I---F---EV---C 490  
Db 4098 VFEDYIYGVTYINNRVFKHKFGHSPLYNLTLGLSHASDVLYVHGHQKQEVTPNCDRKKC 4157  
QY 491 -W---N---N---A---A--- 494  
Db 4158 EWLCLLSPSGVCTCPNGKRLDNGTCVPVPSPTPPDARPPGCTTLQCFNGSGCFLNARR 4217  
QY 495 ---GDK---V---G---AS---A--- 502  
Db 4218 QPKRCQPRYTGDKCBLDQWCEYCHNGGTCAASPSGMPTCRCPGTGTPKRTAQVCAGYC 4277  
QY 503 ---S---DGS---V--- 507  
Db 4278 SNNSTCTVNGNQPOCRCLPGLGDRQCYRQSGPCFENFGTCQMAADGSRQCRCTVYFEG 4337  
QY 508 -C-V---L---DL---R---K 514  
Db 4338 PRCEVKNCSRCLQACVAVNKQTDVTCNCTDGRVAPSLCTCIDHCSNGSGSCTWNSK 4393

## RESULT 9

S02392

alpha-2-macroglobulin receptor precursor - human

N;Alternate names: CD91; LDL receptor-related protein 1; low density lipoprotein receptor  
C;Species: Homo sapiens (man)

C;Date: 14-Aug-1998 #sequence revision 14-Aug-1998 #text\_change 09-Jul-2004

C;Accession: S02392; S30027; I37998; A39210; S12538

R;Herz, J.; Hamann, U.; Rogne, S.; Myklebost, O.; Gausepohl, H.; Stanley, K.K.

EMBO J. 7, 4119-4127, 1988

A;Title: Surface location and high affinity for calcium of a 500-kd liver membrane protein

A;Reference number: S02392; MUID:89210795; PMID:3266596

A;Accession: S02392

A;Status: nucleic acid sequence not shown

A;Molecule type: mRNA

A;Residues: 1-4544 &lt;HER&gt;

A;Cross-references: UNIPROT:Q07954; EMBL:X13916; MID:g34338; PIDN:CAA32112.1; PID:g34339

R;Kristensen, I.

submitted to the EMBL Data Library, October 1990

A;Reference number: S30027



A;Accession: S30027  
A;Molecule type: mRNA  
A;Residues: 3275-3864 <KRI>  
A;Cross-references: EMBL:X55077  
R;Harz, J.; Kowal, R.C.; Goldstein, J.L.; Brown, M.S.  
EMBO J. 9, 1769-1776, 1990  
A;Title: Proteolytic processing of the 600 kd low density lipoprotein receptor-related protein  
A;Reference number: S12538; MUID:90269210; PMID:2112085  
A;Contents: annotation; site of proteolytic cleavage  
R;Kutt, H.; Herz, J.; Stanley, K.K.  
Biochim. Biophys. Acta 1009, 229-236, 1989  
A;Title: Structure of the low-density lipoprotein receptor-related protein (LRP) promotes  
A;Reference number: I37998; MUID:90089395; PMID:2597675  
A;Accession: I37998  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-11 <RES>  
A;Cross-references: EMBL:X15424; NID:G34408; PIDN:CAA33464.1; PID:G34409  
R;Strickland, D.K.; Ashcom, J.D.; Williams, S.; Burgess, W.H.; Migliorini, M.; Argaves,  
J. Biol. Chem. 265, 17401-17404, 1990  
A;Title: Sequence identity between the alpha2-macroglobulin receptor and low density lip  
A;Reference number: A39210; MUID:91009181; PMID:1698775  
A;Accession: A39210  
A;Status: preliminary  
A;Molecule type: protein  
A;Residues: 150-166;234-238,'X',240-245,'X',247-252,'G',686-695;902-916;1096-1109;'S',17  
C;Genetics:  
A;Gene: GDB:LRP1; APR: LRP; A2MR  
A;Cross-references: GDB:111694; OMIM:107770  
A;Map position: 12q13.1-12q13.3  
C;Complex: The alpha-2-macroglobulin receptor complex consists of noncovalently-associated  
ciated protein (see PIR:A39875).  
C;Superfamily: alpha-2-macroglobulin receptor; EGF homology; LDL receptor ligand-binding  
C;Keywords: beta-hydroxyasparagine; beta-hydroxyaspartic acid; calcium binding; glycopro  
F;1-19/Domain: signal sequence #status predicted <SIG>  
F;20-3943/Product: alpha-2-macroglobulin receptor 515K chain #status predicted <515K>  
F;27-64/Domain: LDL receptor ligand-binding repeat homology <LDL1>  
F;72-108/Domain: LDL receptor ligand-binding repeat homology <LDL2>  
F;115-148/Domain: EGF homology <EG1>  
F;154-188/Domain: EGF homology <EG2>  
F;198-239/Domain: LDL receptor WYTD-containing repeat homology <YW01>  
F;240-281/Domain: LDL receptor WYTD-containing repeat homology <YW02>  
F;292-334/Domain: LDL receptor WYTD-containing repeat homology <YW03>  
F;335-378/Domain: LDL receptor WYTD-containing repeat homology <YW04>  
F;379-420/Domain: LDL receptor WYTD-containing repeat homology <YW05>  
F;421-468/Domain: LDL receptor WYTD-containing repeat homology <YW06>  
F;478-519/Domain: EGF homology <EG3>  
F;571-613/Domain: LDL receptor WYTD-containing repeat homology <YW07>  
F;614-659/Domain: LDL receptor WYTD-containing repeat homology <YW08>  
F;660-710/Domain: LDL receptor WYTD-containing repeat homology <YW09>  
F;711-752/Domain: LDL receptor WYTD-containing repeat homology <YW10>  
F;753-799/Domain: LDL receptor WYTD-containing repeat homology <YW11>  
F;807-842/Domain: EGF homology <EG4>  
F;854-890/Domain: LDL receptor ligand-binding repeat homology <LDL3>  
F;895-931/Domain: LDL receptor ligand-binding repeat homology <LDL4>  
F;936-971/Domain: LDL receptor ligand-binding repeat homology <LDL5>  
F;976-1011/Domain: LDL receptor ligand-binding repeat homology <LDL6>  
F;1015-1051/Domain: LDL receptor ligand-binding repeat homology <LDL7>  
F;1062-1097/Domain: LDL receptor ligand-binding repeat homology <LDL8>  
F;1104-1140/Domain: LDL receptor ligand-binding repeat homology <LDL9>  
F;1145-1182/Domain: LDL receptor ligand-binding repeat homology <LDL10>  
F;1185-1221/Domain: EGF homology <EG5>  
F;1227-1261/Domain: EGF homology <EG6>  
F;1269-1308/Domain: LDL receptor WYTD-containing repeat homology <YW12>  
F;1309-1355/Domain: LDL receptor WYTD-containing repeat homology <YW13>  
F;1356-1398/Domain: LDL receptor WYTD-containing repeat homology <YW14>  
F;1399-1445/Domain: LDL receptor WYTD-containing repeat homology <YW15>  
F;1446-1488/Domain: LDL receptor WYTD-containing repeat homology <YW16>  
F;1489-1531/Domain: LDL receptor WYTD-containing repeat homology <YW17>  
F;1540-1578/Domain: EGF homology <EG7>  
F;1583-1626/Domain: LDL receptor WYTD-containing repeat homology <YW18>  
F;1627-1669/Domain: LDL receptor WYTD-containing repeat homology <YW19>  
F;1670-1713/Domain: LDL receptor WYTD-containing repeat homology <YW20>

F;1714-1753/Domain: LDL receptor WYTD-containing repeat homology <YW21>  
F;1754-1796/Domain: LDL receptor WYTD-containing repeat homology <YW22>  
F;1797-1846/Domain: LDL receptor WYTD-containing repeat homology <YW23>  
F;1850-1886/Domain: EGF homology <EG8>  
F;1934-1976/Domain: LDL receptor WYTD-containing repeat homology <YW24>  
F;1977-2019/Domain: LDL receptor WYTD-containing repeat homology <YW25>  
F;2020-2063/Domain: LDL receptor WYTD-containing repeat homology <YW26>  
F;2064-2105/Domain: LDL receptor WYTD-containing repeat homology <YW27>  
F;2106-2151/Domain: LDL receptor WYTD-containing repeat homology <YW28>  
F;2159-2194/Domain: EGF homology <EG9>  
F;2199-2241/Domain: LDL receptor WYTD-containing repeat homology <YW29>  
F;2253-2294/Domain: LDL receptor WYTD-containing repeat homology <YW30>  
F;2344-2388/Domain: LDL receptor WYTD-containing repeat homology <YW31>  
F;2389-2429/Domain: LDL receptor WYTD-containing repeat homology <YW32>  
F;2430-2473/Domain: LDL receptor WYTD-containing repeat homology <YW33>  
F;2482-2517/Domain: EGF homology <EG10>  
F;2524-2561/Domain: LDL receptor ligand-binding repeat homology <LDL1B>  
F;2566-2600/Domain: LDL receptor ligand-binding repeat homology <LDL1C>  
F;2605-2639/Domain: LDL receptor ligand-binding repeat homology <LDL1D>  
F;2652-2688/Domain: LDL receptor ligand-binding repeat homology <LDL1E>  
F;2696-2730/Domain: LDL receptor ligand-binding repeat homology <LDL1F>  
F;2734-2769/Domain: LDL receptor ligand-binding repeat homology <LDL1G>  
F;2774-2812/Domain: LDL receptor ligand-binding repeat homology <LDL1H>  
F;2818-2853/Domain: LDL receptor ligand-binding repeat homology <LDL1I>  
F;2858-2897/Domain: LDL receptor ligand-binding repeat homology <LDL1J>  
F;2904-2939/Domain: LDL receptor ligand-binding repeat homology <LDL1K>  
F;2944-2980/Domain: EGF homology <EG11>  
F;2986-3021/Domain: EGF homology <EG12>  
F;3029-3068/Domain: LDL receptor WYTD-containing repeat homology <YW34>  
F;3069-3113/Domain: LDL receptor WYTD-containing repeat homology <YW35>  
F;3114-3156/Domain: LDL receptor WYTD-containing repeat homology <YW36>  
F;3157-3200/Domain: LDL receptor WYTD-containing repeat homology <YW37>  
F;3201-3241/Domain: LDL receptor WYTD-containing repeat homology <YW38>  
F;3242-3284/Domain: LDL receptor WYTD-containing repeat homology <YW39>  
F;3294-3330/Domain: EGF homology <EG13>  
F;3334-3369/Domain: LDL receptor ligand-binding repeat homology <LDL1L>  
F;3374-3408/Domain: LDL receptor ligand-binding repeat homology <LDL1M>  
F;3413-3448/Domain: LDL receptor ligand-binding repeat homology <LDL1N>  
F;3453-3489/Domain: LDL receptor ligand-binding repeat homology <LDL1O>  
F;3494-3531/Domain: LDL receptor ligand-binding repeat homology <LDL1P>  
F;3536-3570/Domain: LDL receptor ligand-binding repeat homology <LDL1Q>  
F;3575-3609/Domain: LDL receptor ligand-binding repeat homology <LDL1R>  
F;3613-3647/Domain: LDL receptor ligand-binding repeat homology <LDL1S>  
F;3654-3690/Domain: LDL receptor ligand-binding repeat homology <LDL1T>  
F;3695-3731/Domain: LDL receptor ligand-binding repeat homology <LDL1U>  
F;3741-3776/Domain: LDL receptor ligand-binding repeat homology <LDL1V>  
F;3785-3822/Domain: EGF homology <EG14>  
F;3828-3860/Domain: EGF homology <EG15>  
F;3868-3911/Domain: LDL receptor WYTD-containing repeat homology <YW40>  
F;3912-3969/Domain: LDL receptor WYTD-containing repeat homology <YW41>  
F;3944-4544/Product: alpha-2-macroglobulin receptor 85K chain #status predicted <EXT>  
F;3944-4544/Domain: 85K chain extracellular #status predicted <EXT>  
F;3970-4012/Domain: LDL receptor WYTD-containing repeat homology <YW42>  
F;4013-4056/Domain: LDL receptor WYTD-containing repeat homology <YW43>  
F;4057-4099/Domain: LDL receptor WYTD-containing repeat homology <YW44>  
F;4100-4142/Domain: LDL receptor WYTD-containing repeat homology <YW45>  
F;4151-4182/Domain: EGF homology <EG16>  
F;4200-4231/Domain: EGF homology <EG17>  
F;4236-4267/Domain: EGF homology <EG18>  
F;4272-4303/Domain: EGF homology <EG19>  
F;4308-4339/Domain: EGF homology <EG20>  
F;4344-4374/Domain: EGF homology <EG21>  
F;4377-4408/Domain: EGF homology <EG22>  
F;4421-4444/Domain: transmembrane #status predicted <TM>  
F;4445-4544/Domain: intracellular #status predicted <INT>  
F;166,2998/Modified site: erythro-beta-hydroxyasparagine (Asn) #status predicted  
F;2958/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status predicted  
F;4075,4125,4278/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 70.5%; Score 2586.1; DB 1; Length 4544;  
Best Local Similarity 10.0%; Pred. No. 2,2e-52;  
Matches 432; Conservative 65; Mismatches 8; Indels 3809; Gaps 387;

QY 1 MS-----I-----SSDE-----V-----N-F-L----- 11  
Db 88 MSRLCNGVQCDGSDGDEGPHCRLOQNGRSLGQHHCVPTLDGPTCYCNSSFOLQADGKT 147  
QY 12 -----VY-----R-----Y-LQ-----E----- 18  
Db 148 CKDFDECSVYGTCSQLCTNTDGSFICGCVGYLLQPDNRSCKAKNEPDRPPVLLIANQ 207  
QY 19 -----SG-----FS-----H-SAF-T-----F 28  
Db 208 NILATYLSAQVSTIPTSTROTMTAMDFSYANETVCVHVHVGDSAAQTOLKCARMPGLKGF 267  
QY 29 GI-ESH-IS-----Q-----S-N-----I-----NG----- 41  
Db 268 -VDE-HTINISLSLHVQWQAIWLTGNFYFVDDIDRIFVCNVRNGDTCVTLLDLELYNP 325  
QY 42 -----AL-----VP-----P-----A- 47  
Db 326 KGIALDPAMGKVFTDYGOIPKVERCDMDGQNETKLVDSKIIVPHGIGITLDLVSLVYWD 385  
QY 48 A---LI-----SIQKG-L-Q-Y---V-E-----A-EVS-I----- 66  
Db 386 AYLDYIEVDYEGKGTIIQ-GILIEHLYGLTVFENLYATNSDNANAQOKTSVIRVR 444  
QY 67 -N-E-----D-GTL----- 72  
Db 445 FNSTEQVTVTRDKGALHIYHORRPRVRSHACENDQYKPGGCGSDICLLANSHKARTC 504  
QY 73-----F-----DG-----R-----PIE-----SL 81  
Db 505 RCRSGFSLGDKCKPEHELFVYKGRPGIIRGMGMGAKVPDEHMIENLMNPRAL 564  
QY 82 -----S-LI-----D-----AV-M-----PD 90  
Db 565 DFHAETGYFADTTSYLIQKIDGTERITLKGDIHNVGAVDMDGNLYWTDGPK 624  
QY 91 -V-V-----QTR-----Q-----QA-----Y----- 99  
Db 625 KTSVARLEKAAQRTKLIBGKMTHPRAIIVDPLNGHMYTWDWEDPKDSRRGLERAWM 684  
QY 100 -RD-----K-----L-----A-----Q----- 107  
Db 685 DGSHRDIFTVSKTVLWPNGLSLDIPAGRLYVWDAFYDRIETILLNGTDKIVYEGPELNH 744  
QY 108 A-----A-----A-----AA----- 111  
Db 745 AFGLCHHGNYLFWTEYRSGSVYRLERGVGGAPPTVTLRLSERPPIFEIRMYDAQOQVGT 804  
QY 112 -----AA-----A-A-----A-AT-----N-----Q----- 120  
Db 805 NKCRVNGGSSSLCLATPGSRQCAEDQVLDADGVTCLANPSVPPQCPGEFACANS 864  
QY 121 -Q-----G-----S-----A-----K-----N-G----- 127  
Db 865 RCIQERNKCDGNDCLNDSDEAPALCHOCTPSDRFKCENNRCPINRWLDCGNDGCNSE 924  
QY 128 -E-N-T-A-----N-----G-----E-----B----- 135  
Db 925 DESNATCSARTCPFNQFSCASGRCPISWTCDLDDCGDRSDSASCAYTCPFLPTQTC 984  
QY 136 -NG-----A-----H-----T-----I----- 141  
Db 985 NNGRCININWRCDNDCGNSDEAGCSHSCSTQPKCNSGRCPPEHWTCDGNDGCDYS 1044  
QY 142 -----AN-N-----HTD-----M-----M-E-----V- 151  
Db 1045 DETHANCTNQATRPFGCHTDFQCRDLGLCIPLRWRCGDGTDGMDSSDEKSCGVTVC 1104  
QY 152 D-----G-D-----V-----E-I-----PS-----NKAV----- 163  
Db 1105 DPSVKFGKOSARCISKAWCDGNDGNDCEDESNDEENCSLACRPPSHPCANNTSVCLPPDK 1164  
QY 164 V-----L-R-----G-H-----E-----S-----E-----V- 172

Db 1165 LCDGNDGCDGSDGEGELCQDCSLNNGCGSHNCVAPGEGIVCSPGLMELGPDNHTCQIQ 1224  
QY 173 -----F-I-CA-----W-----N----- 178  
Db 1225 SYCAKHLKCSQKCDQNFVKSCCYEGWVLEPDGSCRSRLDPFKFPIIFSNRHRIRIDL 1284  
QY 179 -----P-----V-----S-----D-----LL----- 184  
Db 1285 HKGDYSLVPLGRNTIALDFHLSQALYWDVVDVKIYRGKLLDNGALTSFEVVIQYGLA 1344  
QY 185 -----V-S-----GS-----GD-----S-----T 192  
Db 1345 TPEGLAVDWIAGNIYVWESNLQIEVAKLDGTLRTLLAGDIEHPRAIALPRDGLFWT 1404  
QY 193 -A-RI-----W-N----- 197  
Db 1405 DWDASLPRIEAAASMSGAGRTVHRETGGGWPNGLUTVDYLEKRIILWIDARSDAIYSARYD 1464  
QY 198 -----LS-----E-----NST-----SG----- 205  
Db 1465 GSGHMEVLRGHEFLSHPPFAVTLYGGEVYWDHRTN-TLAKANKWTGHNVTVVQRTNQPF 1523  
QY 206 -----PT-----Q-----LV-L-R-----H-C----- 214  
Db 1524 DLQVYHPSRQPMAPNPCEANGGQPCSHLCLINRYRTVSCACPHLMKLHKDNTTCYEPKK 1583  
QY 215 -----I-----RE----- 217  
Db 1584 FLYARQMEIRGVLDLAPYNYIISFTVPDINVTVDYDAREQRYVSDVTRTOAIKRAF 1643  
QY 218 -G-G-----QDVPSNK-----D-V-----TS-----LD----- 232  
Db 1644 INGTGETVVSADLP-NAHGLAVDWVSRNLFWTSYDTNKKQINVARLDGSKFNAVVQGLE 1702  
QY 233 -----W-----N-S-----EG-TLL-A-TG-----S-Y----- 245  
Db 1703 QPHGLVVLPLRGKLYWTDGDNISMANMDGSRNTLLFSGQKPVGLAIDFPESKLYWISSG 1762  
QY 246 -----D-GF-----A-R-----I-----W-----T-K-DG----- 256  
Db 1763 NHTINCNLDGSGLEVIDAMRSQKATALAIMGDKLWADQVSEKMGCTSKADGSGSVV 1822  
QY 257 -N---LA----- 263  
Db 1823 LRNSTTLVMMKMYDDESIQLDHHKTNPCSVNNGDCSQLCLPTSETTRSCMCTAGYSLRSG 1882  
QY 264 Q-----HKG-----P-----I-F-A----- 271  
Db 1883 QQACEGVGSFLLYSVHEGIRGIPLDPNDKSDALVPVSGTSLAVGIDFHAENDTIYVWDMG 1942  
QY 272 L-----K-----W-----N-----KGN-----P-I-L----- 282  
Db 1943 LSTISRAKRDQTVREDVNTNGIRGVEGIAVDWIAAGNIYWTDOQDFVIEVARLNGSPRYV 2002  
QY 283 -SAGVDK-----T-----T-----I-----I----- 292  
Db 2003 ISQGLDKPRAITVHPKEGYLFWTEWGOYPRIERSLDGTERRVVLNVNWSWPNGISVDYQ 2062  
QY 293 -----W-DA-----H----- 296  
Db 2063 DGKLYWCARDARTKIERIDLETGENREVLSSNNMDMFSVSVFEDFIYWSDRTHANGSIKR 2122  
QY 297 -----TG----- 298  
Db 2123 GSKDNATDSVPLRTGLGVQLKDIKVFNRDRQKSTNVCAVANGCQQLCLYGRGORACAC 2182  
QY 299 -----E-A-----K-----Q-----Q-----F- 304  
Db 2183 AHGMLAEDGASCREYAGLYLISERTILKSITHLSDERNLNAPQFFEDPEHMKNVIALAFD 2242  
QY 305 -----P-----F-H-----S-----A----- 309

Db 2243 YRAGTSGTGNRIFFSDIHFGNIQQINDGSRRTITVENVSGVEGLAYHKGWDTLYWTSY 2302  
QY 310 -----P-AL-----D-----V-D-----W-Q----- 317  
Db 2303 TTSTITHTVDTQTRGAFERETVITMSGDDHPRAFLVDEQNLMTWNEQHPISWRAA 2362  
QY 318 -S-N-----N-----T----- 321  
Db 2363 LSGANVLTIKDIKIRTPNGLAIDHRAEKLIFSADTLKIERCEYDGHSHRYVLKSEBPVHP 2422  
QY 322 -----F-----A-----SC-S-- 326  
Db 2423 FGLAVYGEHIFWTDWVRRAVQRANKHVGSMKLLRVDIPOQPMGIIIAVANDTNSCSPC 2482  
QY 327 -----T-----DM-----CI----- 331  
Db 2483 RINNGGQDLCLLTHQHVNCSCRGRILODDLTCTRAVNSSCRAQDEFECANGECINFSL 2542  
QY 332 -----H-----V-----C-K 335  
Db 2543 TCDGVPHCKDSDEKPSYCNRRCKKTFRQCSNGRCVSNMLWCNMGADDCGDSDEIPCNK 2602  
QY 336 -----F-----Q----- 338  
Db 2603 TACGVGEFRCDGTCTGNSRNCQFVDCEDASDEMNCATDCSSYFRLGVKGLFQPCER 2662  
QY 339 -----D-----RP-----I-----K----- 343  
Db 2663 TSLCYAPSWDCGANDGYSDEDCPGVKRPRCLNYFACPSGRCI PMSWTCDEKDDCE 2722  
QY 344 -----T-----F-----Q-----G-----H----- 348  
Db 2723 HGEDETHCNKFCSEAFQEQNHRCISKQWLCDGSDGDEAAHCEGKTCGPSFSFSCP 2782  
QY 349 -T-----N-----EV-----N-A-I-K----- 356  
Db 2783 GTHVCVPERWLCDDKADGADESAAGCLYNSTCDDREFMQRQCIPKHFVCDHRRD 2842  
QY 357 -----W-----D-----P-T----- 360  
Db 2843 CADGSDSECEYPTCGPSEFRANCGRCLSSRQWECDEGNDCHDQDEAPKNPHCTSPCH 2902  
QY 361 -----G-----N-----L----- 363  
Db 2903 KCWASSQFLCSSGRCAEALLCNGQDDCGDSSDERGCHINECLSRKLSGCSQDCEDLKIG 2962  
QY 364 -----L-----A-----S-----CS-----D----- 369  
Db 2963 FKCRCPGFLKDDGRTCADVDCESTTFPCSQRCINTHGSYKCLCVGYAPRGDPHSCK 3022  
QY 370 -----D-----M-----T-LK-----I-W-----SM 378  
Db 3023 AVTDEPFLLIFANRYLRLKLNLDGSNYLLKQGLNNAVALDPFYREQMYTDTVTQGSN 3082  
QY 379 -----K-----Q-----D-----N-----C----- 383  
Db 3083 IRRMHLNNGSVQVLRHTLSNPDGLAVDWGGLNYCCKGRDTIEVSKLNGAVRTVLVSS 3142  
QY 384 -----V-----H----- 385  
Db 3143 GLREPRALVVDVQNGLYLWTDGHDHSLIGRMGSSRSRVIVDTKITWPNGLTLDYVTER 3202  
QY 386 -----D-----L-Q-----AH----- 390  
Db 3203 IYWADAREDIYEPASLDGSRHVLVSDIPHALTLFEDYVYVTDWETKSNINRAHKTG 3262  
QY 391 -NK-----EI-----Y 395  
Db 3263 TNKTLIISTLHRPMDLHVHALRQPDVPHNPKVNNGGCSNCLLSPGGGHKACACPTNFY 3322  
QY 396 -----T-----IK-W-----S-----P----- 401  
Db 3323 LGSGRTCVSNCTASQFVCKNDKCIPIFWNKCDTDDCGDSDPDPCEPKRPGQFQCS 3382

RESULT 10  
A88852  
protein unc-22 [imported] - Caenorhabditis elegans

QY 402 TG-----P-----G-----TN-----N----- 408  
Db 3383 TGICTNPAFICDGNDCQDNDSEANCDIHVCLPSQFKCTNTNRCIPGIPRCNGODNCGDG 3442  
QY 409 -----PN-----ANL--M----- 414  
Db 3443 EDERDCPEVTCAPNQFQCSITTKRCIPRVVWCDRDNDVDCGSDPANCQTGCVDEPRCK 3502  
QY 415 -----L-A----- 416  
Db 3503 DSGRCIPARWKCDGEDDCGSDGDEPKEDERTCEPYQFRCKNNRCVPGRWQCDYDNDG 3562  
QY 417 -----S-----A-S-----F----- 420  
Db 3563 DNSDEBESCTRPCSESEFSCANGRCIAGRWKCDGDDHDCADGDEKDKTPCPCMDMQFOCKS 3622  
QY 421 -----D-S-----T-VR-----L-W----- 427  
Db 3623 GHCIPLRWRCDADACMDGSDDEACGTGVRTCPDEFQCNNTLCKPLAWKCDGEDDCGDN 3682  
QY 428 -D-----V-----DRGIC-I-----HTLTK 440  
Db 3683 SDENPECARFVCPNRPFRCKNDR-VCLWGRQCDGTNCGDGTDEDECEPPTAHT-T- 3739  
QY 441 H-----Q-----E-----P-----V 445  
Db 3740 HCKDKKEFLRCNRQCLSSSLRCNMFDDCGDSEEDCSIDPKLTSCATNASICGDEARCV 3799  
QY 446 -----Y-----S-----V-----A-F----- 450  
Db 3800 RTEKAAYCACRGFHTVPGQCGDINECLRFQTCSQLCNNTKGGHLCSGARNFMKTHNT 3859  
QY 451 -----S-----P-----D-----CR----- 455  
Db 3860 CKAGSEYQVLYIADNNEIRSLFPFGPHSAYEQAFQGDSESVRIDAMDVHVKAGRVYTNW 3919  
QY 456 -----Y-L-A-S-----G-----S-----F----- 462  
Db 3920 HTGTISYRSLPAAAPPTTSNRHRRQIDRGVTHLNTISGLKMPRGIAIDWAGNVYWTDSGR 3979  
QY 463 D-----K-C-----V-H-I-----W-N-----T- 471  
Db 3980 DVIEVAQMGKGNKTLISGMIDEPHAI VVDPLRGTWYSDWGNHPKPIETAAMDGTLRETL 4039  
QY 472 -----Q-TGAL-V-HS-Y-----R-GT-----G----- 484  
Db 4040 VQDNTQWPTG-LAVDYHNERLYWADAKLSVIGSIRLNGTDPIVAADSCKRGLSHPFSDVF 4098  
QY 485 -----G-----I-F-----EV-----C-W 491  
Db 4099 EDYIYGTVYNNRVFKHKFGHSPVNLVTGGLSHASDVLVLYHQHKQPEVTNPCDRKKCEW 4158  
QY 492 -----N-----A----- 494  
Db 4159 LCLLSPGSPVCTCPNGKRLDNGTCVPVSPPTPPDAPRCTCNLCQFNGSGCFLNARROP 4218  
QY 495 -----GDK-----V-G-A- 500  
Db 4219 KCRCPRYTGDCKELDQCEWHECHRNCGTCAASPSGMPCTCPTGFTGPKCTQQVCAGYCAN 4278  
QY 501 -----S-----ASDGS----- 506  
Db 4279 NSTCTVNGNQPCRCPLPGFLGDRCOYRQCSGYCENFGTCQMAADSGRQRCRTAYFEGSR 4338  
QY 507 --V-----CVL-----DL-----R-----K 514  
Db 4339 CEVKNKSRCLGACVVKQSGDVTCTDGRVAPSLCTCVGHCSNGSGSCTMNSK 4392

C;Species: Caenorhabditis elegans			1503			VVNGKPLKAVGRVXVMSRSGKHIMKIENAVRADSGKFTITLNSSGSCDSTATVTVGPR	1562
C;Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 09-Jul-2004			83			---L-I---D-AVM---PD---VV---Q---T	94
C;Accession: A88852			1563			TPPKGFLDIADVCADGATLSWNPPDDGDBLTGYIVEAQDMNDKGYIEVGKVDPTNT	1622
R;anonymus, The C. elegans Sequencing Consortium.			95			---R---Q---Q---Q---	97
A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biolog			1623			LKNGLRNKGNYKFRVAVNNEGESEPLSADQYTOIKDPWDEPGKPGRIITDFADRID	1682
A;Reference number: A75000; PMID:99069613; PMID:9851916			98			---A---Y---RD---	101
A;Note: see websites genome.wustl.edu/gsc/C.elegans/ and www.sanger.ac.uk/Projects/C_ele			1683			IAMEPPHKGAGAPIEBEYIVEVRDPTKEWKEVRVPTNASISGLKEGKGYQFRVAVNK	1742
A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and			102			---K-LA---	104
A;Accession: A88852			1743			AGPGQSPSEKQAKPFIAPLWKHDLKSIYVAGATVRWEVIGGEPPEVVKFGN	1802
A;Status: preliminary			105			QQ---H---A---A---A---	110
A;Molecule type: DNA			1803			QOLENGIQLTIDTRKNEHTILCIPSAMRSDVGEYRLTVKNSHGADBEKANTLVLDPSKP	1862
A;Residues: 1-6831 <STO>			111			---A---A---A---A---	113
A;Cross-references: UNIPROT:Q23550; GB:chr_IV; PIDN:CAA98081.1; PID:g3881830; GSPDB:GN00			1863			NGPLEVSDVFEDNLNLSWKPDDDDGGEPIEYVEVEKLDATATGRWVPCAKVKDTKAHIDGL	1922
C;Genetics:			114			---A---AA-A---T---N---	119
A;Map position: 4			1923			KGQTVQFRVAVNKEGASDALSTDKTKAKNPYDEPGKTGTPDVVDWDADRVSLWEPP	1982
C;Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;			120			---Q---Q---G---S---A---K---N---G---	127
Query Match 70.5%; Score 2585; DB 2; Length 6831;			1983			KSDGGAPITQYVIEKKGHGRDQECGKVGSGQTNAEILGLKEGBEYQFRVAVNKAQSG	2042
Best Local Similarity 8.0%; Pred. No. 5.4e-52;			128			E---N---T---AN---GE---E---N---136	
Matches 470; Conservative 31; Mismatches 9; Indels 5360; Gaps 386;			2043			EASDPSRKVAVPRNLKPWIDREAMKTIITIKGVNOVEFDPVRGEPKKKEWIFNEKPYD	2102
1 M-----S-----I-----S-S-----D-----6			137			---GA---H---TI-A-N-----143	
784 MTEARRQSLFPGKKVEKDIPLEKTVQQVDKICEWKCTYRPNKIRWYKDRKEIFS			2103			DQKIRIESDYKTRFVLRGATRKAGHAGLYTLTATNASGSKHSVEVILGKSPSPGLPLEV	2162
7 -----EV-----N-----PL-----11			144			N-----H-----TDM-----ME-150	
844 GGLKYKIVIEKNVCTLIINNPEVDDTGKYTEANGVPTHAQLTVLBPPMKYSPLNPLNT			2163			SNVVEDRADLEWKVPDDGGAPIDHYEIEKMDLATGRWVPCGRSETTKTTPVNLQPGHEY	2222
12 --VVR-----Y---LOE-----S-G-----20			151			---V---D-----G-----D-----154	
904 QEYRTRQAVLTCKVNTPRAPLVWYRGSKAIOBQDFRIFIEKDAVGRCITLTIKEVEDDQ			2223			KFRVAVNKEGESDPLTTNTAILAKNPYEPGVKDPKPELVNDKDHVDLAWNAPDDGAP	2282
21 -----FS-----HSAF-----T-----27			155			VE---I---P-----S---NKA-----162	
964 AEWARTIQDFVKQVYVEEPRT-FVVPMSQKVNESDLATLETVDNDKDAEVVWHD			2283			IEAFVIEKKDKNGRWEALVWPGDKTATVPNLKEGEEYQFRISARNKAGTGDPSDR	2342
28 -----F-----G-----IESH-----33			163			VV---L---R---G---H-----ES-----170	
1023 GKRIDIGVKFVBSNRKRLLIINGARIEDHGEYKCTTKDDRTMAQLIVDAKNKFIVAL			2343			VVAKPRNLAPRIHREDLSDTTVKVGATLKFIVHIDGEPADVTWSFNGKIGESKAQIEN	2402
34 -----IS-----QS-N-----I--N 40			171			---F-----I---C-----175	
1083 KOTEVIEKDDVTLMCQTKTPTGPIWFRNGKQISSMGGKFETQSRNGTHTLKIGKLENN			2463			TLDWKPPDDGGEPIEFYIEIEKMTKDIWVPCGRSGDTHTVDSLNGKGDHYKFRVAVN	2522
41 -----G-----A-----L-----VP-----P 46			176			---A-----A-----176	
1143 EADVVEIDQAGLRGSCNVTVLEAEKRPILNWKPKKIEAKAGPCVVKVPFIKGTGRGDP			2523			SEGSPDPLETETDILAKNPFDRDRGRPREPTDWDSDHVDLKWDPPLSDGAPIEEYQIE	2582
47 -AA-L-----IS-----Q-----54			177			---W-----177	
1203 KAQILKNGKPIDEEMRKLVEIHKDDVAEIVFKNPOLADTGKWALELGNAGTALAPFEL			2583			KRTKYGRWEPAITVPGGQTATTVPDLTPNEEYBFRVAVNKGKGGPSDPSDASKAVIAKPRN	2642
55 -----KG-LQ-----YV---E---A-----62							
1263 FVKDKPKPGPLETKVNTAEGLDLVWGTTPDPPDGGAPKAVIIEWQGRSGNNAKVGETK							
63 -----E---V-----SI-----N-----67							
1323 GTDFKVKDLKEHGEYKFRVKNALNEGLSDPLTGSLAKNPYGVGPKPKNDADIVDKDH							
68 -----EG-----T-----L-----72							
1383 CTLAWEPPEDGGAPITGYIIIEREKSEKQHWQGVQTKPDCCCLTDKKVEDKEYLYRVK							
73 -----F-----D-----74							
1443 AVNKAQPGDPCDHGKPIKWKAKASPEFTGGIKDLRLKVGETIKYDVPISGEPLPECLW							
75 ---GRP-----IE-----S-----L-----S-----82							

178 QY -----N-----pvs-----D-----L----- 183  
2643 Db LKPHIDRALKNLTIRKAGQISFDVPVSGEPAPVTWHDNREINGRVLNDEYQS 2702  
184 QY -----LV-----SG-----S-----GD----- 190  
2703 Db KLVVQWEMRGDSGFTTIKAVNAGEDATVKINVIDKPTSPNGPLDVSVDVHGHTLNWR 2762  
191 QY -----S-----TA-----R----- 194  
2763 Db APDDGGIPIENYVIEKYDTASGRWPAKAVAGDKTTAVVDGLIPGHEYKFRVAANAEG 2822  
195 QY -----I----- 195  
2823 Db ESDPLEFTGTTAKDPDKPKGTNAPEITDWDKHDVLEWKPPANDGGAPIBEYVVEMKD 2882  
196 QY -----W-----NL-----S-----E----- 200  
2883 Db EFSFNDVAHVAPAGOTNATVGNLKEGSKYEFIRAKNKAGLGDPSDSASAVAKARNVPP 2942  
201 QY -----NS-----TSG-----PT-----Q-----L-----V 210  
2943 Db VIDRNSIQEIKVKAGQFSLNIPVSGEPTTITWTPEGTPEVSDRKLNNEDGKTKFHV 3002  
211 QY -----LR-----H-----C----- 214  
3003 Db KRALRSDTGTYYIIKAENENGDTAEVKVTVLDHPSPRGLDVTNIVKDCDLAWKEPED 3062  
215 QY -----I----- 215  
3063 Db DGAIEISHVYIEKQDAATGRWTACGESKOTNFHVDLITQGEYKFRVKA VNRHGDSDPLE 3122  
216 QY -----RE----- 217  
3123 Db AREALIAKOPFRADKPGTPEIVDKDHADLKWTPPADGGAPIBEGYLVMETPSGDWV 3182  
218 QY -----G----- 218  
3183 Db PAVTVAGELTATVDGLKPGQTYQFRVKALNKAGESTPSPSRTMVAKPRHLPKINRDM 3242  
219 QY -----GO-----DV-----P----- 223  
3243 Db FVAQVRKAGQTLNFDVNVGEPAKIEWFLNGSLSSGGNTHIDNNTDNNTKLTTKSTAR 3302  
224 QY -----S-----N-----KDV-----S-----LD----- 232  
3303 Db ADSGKYIVATNESGKDEHVDVNIILDIPGAEPGLRHKIDITKESVVLKWDPLEDDGGSP 3362  
233 QY -----W-----N-----SS-----GT-----L----- 239  
3363 Db ITNYVVEKQEDGGRWVPCGETSPTSLKVNKLSGHEYKFRVKA VNRQGTSAPLTSDHAIV 3422  
240 QY -----L-----A-----T-----G-----SY-----D 246  
3423 Db AKNPFDEPDAPTDTVPVMDKHDHVLWKPPANDGGAPIDAYIVEKKDKFGDWVECARVD 3482  
247 QY -----G-----F-----A-----RI 251  
3483 Db GKTTKATADNLTPGETYQFRVKA VNRKAGPKPSDPTGNVVAKPRRMAPKLNLAGLDR 3542  
252 QY -----W-----TKD-----G-----NL-----A-----S-----T-----L----- 262  
3543 Db KAGTPIKLDIAPEGEPAKVAWKANDATIDTGARADVTTPTSSAIIHIFSAVRGDTGVYK 3602  
263 QY -----G-----Q-----HK-----G-----P-----I 269  
3603 Db IIVENEHGKDTACQCNVTLVDVPGTPEGPLKIDRIHKEGCTLNWKPPPTDNGGTDVLHYVE 3662  
270 QY -----F-----A-----L----- 272  
3663 Db KMDTSRGTWQEVGTFDCTAKVNKLVPGEYAFRVRKAVNLOGESKPLEAEPIIAKNQPD 3722

273 QY -----KWN-----KK-----G----- 278  
3723 Db VPDVDKPEVTDWDKDRIDIKNPTANNNGCAPVTGYIVEKKEKGSALWTEAGTPTGTFPS 3782  
279 QY -----N-----P-----IL-----S-----AG----- 285  
3783 Db ADNLKPGVEYFRVIAVNAAGSPSDPTDQITKARYLKPKILTASRKIKIKAGFTHNL 3842  
286 QY -----VD-----K-----TT----- 290  
3843 Db EVDFIGAPDPTATWTVDGSAALAPPELLVDAGSSTTSIFPPSAKRADSGNYKLVKKNELG 3902  
291 QY -----I-----I-----W-----D----- 294  
3903 Db EDEAIFEVIVQDRPSAPEGPLEVSDVTKDSCVLNWKPPKDDGAEISNYVVEKRDTKTNT 3962  
295 QY -----A-----H-----T-----G-----E-----AKOOF----- 304  
3963 Db WVPVSFAVTGTSITVPKLTGEGHEYFRVMAENTFGRSDSLNTDPEVLAKDPFGTGGPKGR 4022  
305 QY -----P-----F----- 306  
4023 Db PEIVDTDNHDHIDIKWDPDRDNGGSPVDHYDIERKDAKTGRWIKWNTSPVOGTAFSDTRVQ 4082  
307 QY -----H-----SA-----P-----A-----LD-----V----- 314  
4083 Db KGHTEYRVVAVNKAQPGQPSDSSAAATAKPMHEAPKFDLDLDGKEFRVKAGEPLVITIP 4142  
315 QY -----D-----W-----Q-----S----- 318  
4143 Db FTASPOQDISWTEGKGLAGVETDTSQTKLVIPSTRSDSGPVKIKAVNPYGEAEANIK 4202  
319 QY -----N-----N----- 320  
4203 Db ITVIDKGAPEKITYPVAVSRHTCTLNWDAPKDDGAEIAGYKIEYEVSGSIQWIKVPLGI 4262  
321 QY -----T-----F-----A-----S-----C-----ST----- 327  
4263 Db SGTAYTVRGLEHQOQYFRFIRAEANAVGLSDYCOQVPVWIKDPPDPGAPSTPEITGYDTN 4322  
328 QY -----H-----V-----C-----D-----M-----CI----- 331  
4323 Db QVSLAWNPRDDGGSPILGYVVERFEKGGGWAPVKMPVWKGTECIVPLGHENETVQFR 4382  
332 QY -----H-----V-----C----- 334  
4383 Db VRVNAAGHGEPSNGSEBPVTCRPYVEKPGAPDAPRVGKITKNSAELTWNRLRDLGGAPID 4442  
335 QY -----KLK----- 337  
4443 Db GYIVEKKLGDNDWTRCNDKPVDRDTAFVKNLGEKEEYFRVIAVNSAGEGESKPSDLV 4502  
338 QY -----Q-----D-----R-----PI----- 342  
4503 Db LIEEQGRPIFDINNLKIDTVRAGETIQIRIPVAGGNPKPIIDLFGNSPIFENERTVVD 4562  
343 QY -----K-----T-----F-----Q----- 346  
4563 Db VNPGEIVITTTGSKRSDAGPYKISATNKYKQDCKLVNPFVLDAPGKPTGPIRATDQADA 4622  
347 QY -----GH-----T-----N-----E-----V-----N 353  
4623 Db MTLWRPXPKNCGDAITNYVVEKRTFPGDWVTVGHVGTTLVRNLDANTPPEFRVAEN 4682  
354 QY -----AI-----K-----W-----D-----P----- 359  
4683 Db QYGVGEPLTDDAI VAKNPFDTFGAPQPEAVETSEAITLQWTRTPSDGGAPIQGVVIE 4742  
360 QY -----T-----GN----- 362  
4743 Db KREVGSTWTKAAGFNILDKHRVTGLTPKKTYEFRVAAVNAAGQGEYSVNSVPIITADNA 4802  
363 QY -----L----- 363



A;Residues: 792-6839 <BEN2>  
A;Cross-references: EMBL:X15423; NID:g6897; PIDN:CAA33463.1; PID:g6898  
A;Experimental source: var. Bristol  
R;Benian, G.M.; Kiff, J.E.; Neckelmann, N.; Moerman, D.G.; Waterston, R.H.  
Nature 342, 45-50, 1989  
A;Title: Sequence of an unusually large protein implicated in regulation of myosin activity  
A;Reference number: S06797; MUID:90044042; PMID:2812002  
A;Accession: S06797  
A;Status: nucleic acid sequence not shown  
A;Molecule type: DNA  
A;Residues: 806-1175, 1178-1998, 'Y', 2000-3040, 'I', 3042-3335, 'I', 3337-5693, 5696-6359, 'I', 6361-6362, 6364-6365, 6366-6367, 6368-6369, 6370-6371, 6372-6373, 6374-6375, 6376-6377, 6378-6379, 6380-6381, 6382-6383, 6384-6385, 6386-6387, 6388-6389, 6390-6391, 6392-6393, 6394-6395, 6396-6397, 6398-6399, 6400-6401, 6402-6403, 6404-6405, 6406-6407, 6408-6409, 6410-6411, 6412-6413, 6414-6415, 6416-6417, 6418-6419, 6420-6421, 6422-6423, 6424-6425, 6426-6427, 6428-6429, 6430-6431, 6432-6433, 6434-6435, 6436-6437, 6438-6439, 6440-6441, 6442-6443, 6444-6445, 6446-6447, 6448-6449, 6450-6451, 6452-6453, 6454-6455, 6456-6457, 6458-6459, 6460-6461, 6462-6463, 6464-6465, 6466-6467, 6468-6469, 6470-6471, 6472-6473, 6474-6475, 6476-6477, 6478-6479, 6480-6481, 6482-6483, 6484-6485, 6486-6487, 6488-6489, 6490-6491, 6492-6493, 6494-6495, 6496-6497, 6498-6499, 6500-6501, 6502-6503, 6504-6505, 6506-6507, 6508-6509, 6510-6511, 6512-6513, 6514-6515, 6516-6517, 6518-6519, 6520-6521, 6522-6523, 6524-6525, 6526-6527, 6528-6529, 6530-6531, 6532-6533, 6534-6535, 6536-6537, 6538-6539, 6540-6541, 6542-6543, 6544-6545, 6546-6547, 6548-6549, 6550-6551, 6552-6553, 6554-6555, 6556-6557, 6558-6559, 6560-6561, 6562-6563, 6564-6565, 6566-6567, 6568-6569, 6570-6571, 6572-6573, 6574-6575, 6576-6577, 6578-6579, 6580-6581, 6582-6583, 6584-6585, 6586-6587, 6588-6589, 6590-6591, 6592-6593, 6594-6595, 6596-6597, 6598-6599, 6600-6601, 6602-6603, 6604-6605, 6606-6607, 6608-6609, 6610-6611, 6612-6613, 6614-6615, 6616-6617, 6618-6619, 6620-6621, 6622-6623, 6624-6625, 6626-6627, 6628-6629, 6630-6631, 6632-6633, 6634-6635, 6636-6637, 6638-6639, 6640-6641, 6642-6643, 6644-6645, 6646-6647, 6648-6649, 6650-6651, 6652-6653, 6654-6655, 6656-6657, 6658-6659, 6660-6661, 6662-6663, 6664-6665, 6666-6667, 6668-6669, 6670-6671, 6672-6673, 6674-6675, 6676-6677, 6678-6679, 6680-6681, 6682-6683, 6684-6685, 6686-6687, 6688-6689, 6690-6691, 6692-6693, 6694-6695, 6696-6697, 6698-6699, 6700-6701, 6702-6703, 6704-6705, 6706-6707, 6708-6709, 6710-6711, 6712-6713, 6714-6715, 6716-6717, 6718-6719, 6720-6721, 6722-6723, 6724-6725, 6726-6727, 6728-6729, 6730-6731, 6732-6733, 6734-6735, 6736-6737, 6738-6739, 6740-6741, 6742-6743, 6744-6745, 6746-6747, 6748-6749, 6750-6751, 6752-6753, 6754-6755, 6756-6757, 6758-6759, 6760-6761, 6762-6763, 6764-6765, 6766-6767, 6768-6769, 6770-6771, 6772-6773, 6774-6775, 6776-6777, 6778-6779, 6780-6781, 6782-6783, 6784-6785, 6786-6787, 6788-6789, 6790-6791, 6792-6793, 6794-6795, 6796-6797, 6798-6799, 6800-6801, 6802-6803, 6804-6805, 6806-6807, 6808-6809, 6810-6811, 6812-6813, 6814-6815, 6816-6817, 6818-6819, 6820-6821, 6822-6823, 6824-6825, 6826-6827, 6828-6829, 6830-6831, 6832-6833, 6834-6835, 6836-6837, 6838-6839, 6840-6841, 6842-6843, 6844-6845, 6846-6847, 6848-6849, 6850-6851, 6852-6853, 6854-6855, 6856-6857, 6858-6859, 6860-6861, 6862-6863, 6864-6865, 6866-6867, 6868-6869, 6870-6871, 6872-6873, 6874-6875, 6876-6877, 6878-6879, 6880-6881, 6882-6883, 6884-6885, 6886-6887, 6888-6889, 6890-6891, 6892-6893, 6894-6895, 6896-6897, 6898-6899, 6900-6901, 6902-6903, 6904-6905, 6906-6907, 6908-6909, 6910-6911, 6912-6913, 6914-6915, 6916-6917, 6918-6919, 6920-6921, 6922-6923, 6924-6925, 6926-6927, 6928-6929, 6930-6931, 6932-6933, 6934-6935, 6936-6937, 6938-6939, 6940-6941, 6942-6943, 6944-6945, 6946-6947, 6948-6949, 6950-6951, 6952-6953, 6954-6955, 6956-6957, 6958-6959, 6960-6961, 6962-6963, 6964-6965, 6966-6967, 6968-6969, 6970-6971, 6972-6973, 6974-6975, 6976-6977, 6978-6979, 6980-6981, 6982-6983, 6984-6985, 6986-6987, 6988-6989, 6990-6991, 6992-6993, 6994-6995, 6996-6997, 6998-6999, 7000-7001, 7002-7003, 7004-7005, 7006-7007, 7008-7009, 7010-7011, 7012-7013, 7014-7015, 7016-7017, 7018-7019, 7020-7021, 7022-7023, 7024-7025, 7026-7027, 7028-7029, 7030-7031, 7032-7033, 7034-7035, 7036-7037, 7038-7039, 7040-7041, 7042-7043, 7044-7045, 7046-7047, 7048-7049, 7050-7051, 7052-7053, 7054-7055, 7056-7057, 7058-7059, 7060-7061, 7062-7063, 7064-7065, 7066-7067, 7068-7069, 7070-7071, 7072-7073, 7074-7075, 7076-7077, 7078-7079, 7080-7081, 7082-7083, 7084-7085, 7086-7087, 7088-7089, 7090-7091, 7092-7093, 7094-7095, 7096-7097, 7098-7099, 7100-7101, 7102-7103, 7104-7105, 7106-7107, 7108-7109, 7110-7111, 7112-7113, 7114-7115, 7116-7117, 7118-7119, 7120-7121, 7122-7123, 7124-7125, 7126-7127, 7128-7129, 7130-7131, 7132-7133, 7134-7135, 7136-7137, 7138-7139, 7140-7141, 7142-7143, 7144-7145, 7146-7147, 7148-7149, 7150-7151, 7152-7153, 7154-7155, 7156-7157, 7158-7159, 7160-7161, 7162-7163, 7164-7165, 7166-7167, 7168-7169, 7170-7171, 7172-7173, 7174-7175, 7176-7177, 7178-7179, 7180-7181, 7182-7183, 7184-7185, 7186-7187, 7188-7189, 7190-7191, 7192-7193, 7194-7195, 7196-7197, 7198-7199, 7200-7201, 7202-7203, 7204-7205, 7206-7207, 7208-7209, 7210-7211, 7212-7213, 7214-7215, 7216-7217, 7218-7219, 7220-7221, 7222-7223, 7224-7225, 7226-7227, 7228-7229, 7230-7231, 7232-7233, 7234-7235, 7236-7237, 7238-7239, 7240-7241, 7242-7243, 7244-7245, 7246-7247, 7248-7249, 7250-7251, 7252-7253, 7254-7255, 7256-7257, 7258-7259, 7260-7261, 7262-7263, 7264-7265, 7266-7267, 7268-7269, 7270-7271, 7272-7273, 7274-7275, 7276-7277, 7278-7279, 7280-7281, 7282-7283, 7284-7285, 7286-7287, 7288-7289, 7290-7291, 7292-7293, 7294-7295, 7296-7297, 7298-7299, 7300-7301, 7302-7303, 7304-7305, 7306-7307, 7308-7309, 7310-7311, 7312-7313, 7314-7315, 7316-7317, 7318-7319, 7320-7321, 7322-7323, 7324-7325, 7326-7327, 7328-7329, 7330-7331, 7332-7333, 7334-7335, 7336-7337, 7338-7339, 7340-7341, 7342-7343, 7344-7345, 7346-7347, 7348-7349, 7350-7351, 7352-7353, 7354-7355, 7356-7357, 7358-7359, 7360-7361, 7362-7363, 7364-7365, 7366-7367, 7368-7369, 7370-7371, 7372-7373, 7374-7375, 7376-7377, 7378-7379, 7380-7381, 7382-7383, 7384-7385, 7386-7387, 7388-7389, 7390-7391, 7392-7393, 7394-7395, 7396-7397, 7398-7399, 7400-7401, 7402-7403, 7404-7405, 7406-7407, 7408-7409, 7410-7411, 7412-7413, 7414-7415, 7416-7417, 7418-7419, 7420-7421, 7422-7423, 7424-7425, 7426-7427, 7428-7429, 7430-7431, 7432-7433, 7434-7435, 7436-7437, 7438-7439, 7440-7441, 7442-7443, 7444-7445, 7446-7447, 7448-7449, 7450-7451, 7452-7453, 7454-7455, 7456-7457, 7458-7459, 7460-7461, 7462-7463, 7464-7465, 7466-7467, 7468-7469, 7470-7471, 7472-7473, 7474-7475, 7476-7477, 7478-7479, 7480-7481, 7482-7483, 7484-7485, 7486-7487, 7488-7489, 7490-7491, 7492-7493, 7494-7495, 7496-7497, 7498-7499, 7500-7501, 7502-7503, 7504-7505, 7506-7507, 7508-7509, 7510-7511, 7512-7513, 7514-7515, 7516-7517, 7518-7519, 7520-7521, 7522-7523, 7524-7525, 7526-7527, 7528-7529, 7530-7531, 7532-7533, 7534-7535, 7536-7537, 7538-7539, 7540-7541, 7542-7543, 7544-7545, 7546-7547, 7548-7549, 7550-7551, 7552-7553, 7554-7555, 7556-7557, 7558-7559, 7560-7561, 7562-7563, 7564-7565, 7566-7567, 7568-7569, 7570-7571, 7572-7573, 7574-7575, 7576-7577, 7578-7579, 7580-7581, 7582-7583, 7584-7585, 7586-7587, 7588-7589, 7590-7591, 7592-7593, 7594-7595, 7596-7597, 7598-7599, 7600-7601, 7602-7603, 7604-7605, 7606-7607, 7608-7609, 7610-7611, 7612-7613, 7614-7615, 7616-7617, 7618-7619, 7620-7621, 7622-7623, 7624-7625, 7626-7627, 7628-7629, 7630-7631, 7632-7633, 7634-7635, 7636-7637, 7638-7639, 7640-7641, 7642-7643, 7644-7645, 7646-7647, 7648-7649, 7650-7651, 7652-7653, 7654-7655, 7656-7657, 7658-7659, 7660-7661, 7662-7663, 7664-7665, 7666-7667, 7668-7669, 7670-7671, 7672-7673, 7674-7675, 7676-7677, 7678-7679, 7680-7681, 7682-7683, 7684-7685, 7686-7687, 7688-7689, 7690-7691, 7692-7693, 7694-7695, 7696-7697, 7698-7699, 7700-7701, 7702-7703, 7704-7705, 7706-7707, 7708-7709, 7710-7711, 7712-7713, 7714-7715, 7716-7717, 7718-7719, 7720-7721, 7722-7723, 7724-7725, 7726-7727, 7728-7729, 7730-7731, 7732-7733, 7734-7735, 7736-7737, 7738-7739, 7740-7741, 7742-7743, 7744-7745, 7746-7747, 7748-7749, 7750-7751, 7752-7753, 7754-7755, 7756-7757, 7758-7759, 7760-7761, 7762-7763, 7764-7765, 7766-7767, 7768-7769, 7770-7771, 7772-7773, 7774-7775, 7776-7777, 7778-7779, 7780-7781, 7782-7783, 7784-7785, 7786-7787, 7788-7789, 7790-7791, 7792-7793, 7794-7795, 7796-7797, 7798-7799, 7800-7801, 7802-7803, 7804-7805, 7806-7807, 7808-7809, 7810-7811, 7812-7813, 7814-7815, 7816-7817, 7818-7819, 7820-7821, 7822-7823, 7824-7825, 7826-7827, 7828-7829, 7830-7831, 7832-7833, 7834-7835, 7836-7837, 7838-7839, 7840-7841, 7842-7843, 7844-7845, 7846-7847, 7848-7849, 7850-7851, 7852-7853, 7854-7855, 7856-7857, 7858-7859, 7860-7861, 7862-7863, 7864-7865, 7866-7867, 7868-7869, 7870-7871, 7872-7873, 7874-7875, 7876-7877, 7878-7879, 7880-7881, 7882-7883, 7884-7885, 7886-7887, 7888-7889, 7890-7891, 7892-7893, 7894-7895, 7896-7897, 7898-7899, 7900-7901, 7902-7903, 7904-7905, 7906-7907, 7908-7909, 7910-7911, 7912-7913, 7914-7915, 7916-7917, 7918-7919, 7920-7921, 7922-7923, 7924-7925, 7926-7927, 7928-7929, 7930-7931, 7932-7933, 7934-7935, 7936-7937, 7938-7939, 7940-7941, 7942-7943, 7944-7945, 7946-7947, 7948-7949, 7950-7951, 7952-7953, 7954-7955, 7956-7957, 7958-7959, 7960-7961, 7962-7963, 7964-7965, 7966-7967, 7968-7969, 7970-7971, 7972-7973, 7974-7975, 7976-7977, 7978-7979, 7980-7981, 7982-7983, 7984-7985, 7986-7987, 7988-7989, 7990-7991, 7992-7993, 7994-7995, 7996-7997, 7998-7999, 8000-8001, 8002-8003, 8004-8005, 8006-8007, 8008-8009, 8010-8011, 8012-8013, 8014-8015, 8016-8017, 8018-8019, 8020-8021, 8022-8023, 8024-8025, 8026-8027, 8028-8029, 8030-8031, 8032-8033, 8034-8035, 8036-8037, 8038-8039, 8040-8041, 8042-8043, 8044-8045, 8046-8047, 8048-8049, 8050-8051, 8052-8053, 8054-8055, 8056-8057, 8058-8059, 8060-8061, 8062-8063, 8064-8065, 8066-8067, 8068-8069, 8070-8071, 8072-8073, 8074-8075, 8076-8077, 8078-8079, 8080-8081, 8082-8083, 8084-8085, 8086-8087, 8088-8089, 8090-8091, 8092-8093, 8094-8095, 8096-8097, 8098-8099, 8100-8101, 8102-8103, 8104-8105, 8106-8107, 8108-8109, 8110-8111, 8112-8113, 8114-8115, 8116-8117, 8118-8119, 8120-8121, 8122-8123, 8124-8125, 8126-8127, 8128-8129, 8130-8131, 8132-8133, 8134-8135, 8136-8137, 8138-8139, 8140-8141, 8142-8143, 8144-8145, 8146-8147, 8148-8149, 8150-8151, 8152-8153, 8154-8155, 8156-8157, 8158-8159, 8160-8161, 8162-8163, 8164-8165, 8166-8167, 8168-8169, 8170-8171, 8172-8173, 8174-8175, 8176-8177, 8178-8179, 8180-8181, 8182-8183, 8184-8185, 8186-8187, 8188-8189, 8190-8191, 8192-8193, 8194-8195, 8196-8197, 8198-8199, 8200-8201, 8202-8203, 8204-8205, 8206-8207, 8208-8209, 8210-8211, 8212-8213, 8214-8215, 8216-8217, 8218-8219, 8220-8221, 8222-8223, 8224-8225, 8226-8227, 8228-8229, 8230-8231, 8232-8233, 8234-8235, 8236-8237, 8238-8239, 8240-8241, 8242-8243, 8244-8245, 8246-8247, 8248-8249, 8250-8251, 8252-8253, 8254-8255, 8256-8257, 8258-8259, 8260-8261, 8262-8263, 8264-8265, 8266-8267, 8268-8269, 8270-8271, 8272-8273, 8274-8275, 8276-8277, 8278-8279, 8280-8281, 8282-8283, 8284-8285, 8286-8287, 8288-8289, 8290-8291, 8292-8293, 8294-8295, 8296-8297, 8298-8299, 8300-8301, 8302-8303, 8304-8305, 8306-8307, 8308-8309, 8310-8311, 8312-8313, 8314-8315, 8316-8317, 8318-8319, 8320-8321, 8322-8323, 8324-8325, 8326-8327, 8328-8329, 8330-8331, 8332-8333, 8334-8335, 8336-8337, 8338-8339, 8340-8341, 8342-8343, 8344-8345, 8346-8347, 8348-8349, 8350-8351, 8352-8353, 8354-8355, 8356-8357, 8358-8359, 8360-8361, 8362-8363, 8364-8365, 8366-8367, 8368-8369, 8370-8371, 8372-8373, 8374-8375, 8376-8377, 8378-8379, 8380-8381, 8382-8383, 8384-8385, 8386-8387, 8388-8389, 8390-8391, 8392-8393, 8394-8395, 8396-8397, 8398-8399, 8400-8401, 8402-8403, 8404-8405, 8406-8407, 8408-8409, 8410-8411, 8412-8413, 8414-8415, 8416-8417, 8418-8419, 8420-8421, 8422-8423, 8424-8425, 8426-8427, 8428-8429, 8430-8431, 8432-8433, 8434-8435, 8436-8437, 8438-8439, 8440-8441, 8442-8443, 8444-8445, 8446-8447, 8448-8449, 8450-8451, 8452-8453, 8454-8455, 8456-8457, 8458-8459, 8460-8461, 8462-8463, 8464-8465, 8466-8467, 8468-8469, 8470-8471, 8472-8473, 8474-8475, 8476-8477, 8478-8479, 8480-8481, 8482-8483, 8484-8485, 8486-8487, 8488-8489, 8490-8491, 8492-8493, 8494-8495, 8496-8497, 8498-8499, 8500-8501, 8502-8503, 8504-8505, 8506-8507, 8508-8509, 8510-8511, 8512-8513, 8514-8515, 8516-8517, 8518-8519, 8520-8521, 8522-8523, 8524-8525, 8526-8527, 8528-8529,

QY 137 -----GA-H-----TI-A-N-----143  
Db 2111 DQKRIESEDYKTRFVLGRATKXAGLYTLTATNASGDKHSEVIVLKGKSPSLGLEV 2170  
QY 144 N-----H-----TDM-----ME-150  
Db 2171 SNVYEDRADLEWKVPEDDGGAPIDHIEIEKMDLATGRWVPCGRSETTKTVPNLQGHY 2230  
QY 151 -----V-----D-----G-----D-----154  
Db 2231 KFRVRVKNKEGEDPLTTNTAILAKNPEYGVKDPKPELVDMKDHVDLAWNAPDDGGAP 2290  
QY 155 VE-I-----P-----S-NKA-----162  
Db 2291 IEAFVIEKKDKNGRWEALVVPDQKTATVPNLKEGEYQFRISARNKAGTGDSPSPDR 2350  
QY 163 VV-L-R-----G-----H-----ES-----170  
Db 2351 VVAKRNLAPRIHREDLSDTTVKVGATLKFIVHIDGEPADPVTWSPNGKIGESKAQIEN 2410  
QY 171 -----EV-----172  
Db 2411 EPYISRALPKLRKQSGKTYTITATNINGTDSVTINIKVSKPKPKPIEVTDPEDRA 2470  
QY 173 -----F-----I-C-----175  
Db 2471 TLDWKPPDDGGEPIEFYIEKMKTKDGIWVPCGRSGDTHFTVDSLNGKDHYKFRKAVN 2530  
QY 176 -----A-----176  
Db 2531 SEGSPDLETETDILAKNPFDRDRGRPEPTDWDSDHVDLKWDPPLSDGAPIEEYQIE 2590  
QY 177 -----W-----177  
Db 2591 KRTKYGRWEPAITVPGGQTATVPDLTPNEEYEFVRVAVNKGSPSPSDASKAVIAKPRN 2650  
QY 178 -----N-----PVS-----D-----L-----183  
Db 2651 LKPHIDRALKNLTIKAGQISFDPVPSGEPAPVTWHPDNREIRNGRVRKLDNPEYQS 2710  
QY 184 LV-SG-----S-GD-----190  
Db 2711 KLVVKQMERGDSGFTTIKAVNANGEDATVKINVIDKPTSPNGPLDVSVDVHGDHVTLNR 2770  
QY 191 -----S-----TA-----R-----194  
Db 2771 APDDGGIPIENYVIEKYDTASGRWVPAKVAGDKTTAVVDGLIPGHEYKFRVAANAE 2830  
QY 195 -----I-----195  
Db 2831 ESDPLEFTGTLKDPFKGTNAPEITDWDKDHVDLEWKPPANDGGAPIEEYVEMKD 2890  
QY 196 -----W-----NL-S-E-----200  
Db 2891 EFSFNDVAHPAGQTNATVGNLKEGSKYEFRIKAKNAGLGDPSDSASAVAKARNVPP 2950  
QY 201 -----NS-----TSG-PT-----Q-L-----V 210  
Db 2951 VIDRNSIQEIKVKAGQDFSLNIPVSGEPTTITWTFTGTPVESDDRMKLANNEDGKTFHV 3010  
QY 211 -----LR-----H-----C-----214  
Db 3011 KRALRSDTGTVIIKAENENGDTAEVKVTVLDFHSSPRGLDVTNIVKDGCDLAWKEPED 3070  
QY 215 -----I-----215  
Db 3071 DGGAEISHYVIEKQDAATGRWTACGESKDTNFHVDLDTQGHEYKFRKAVNRHGDSDPLE 3130  
QY 216 RE-----217  
Db 3131 AREAIIAKDPFDRADKPGTPEIVDWDKDHADLKWTTPADGGAPIEGYLVEMRTPSGDWV 3190

QY 218 -----G-----218  
Db 3191 PAVTVGAGELTATVDGLKPGQTYQFRVKALNKAQSTPSDPSRTWVAKPRHLAPKINRDM 3250  
QY 219 -----GO-DV-----P-----223  
Db 3251 FVAQRVKAGQTLNFDVNVEGEPAPKIEWFLNGSLSSGNTNHNNTDNNNTKLTTKSTAR 3310  
QY 224 -----S-----N-----KDV-T-S-----LD-----232  
Db 3311 ADSGKYIVATNESKDBHEVDVNILDIPGAPEGLRHKDIITKESVVLKWDPELDGGSF 3370  
QY 233 -----W-----N-SE-----GT-L-----239  
Db 3371 ITNYVVEKQEDGGRVWVPCGETSDTSIKVNLSEGHYKFRKAVNRQTSAPLTSDBAIV 3430  
QY 240 -----L-A-T-----G-SY-----D 246  
Db 3431 AKNPPDEPDAPDPTVDWDKDHVDLEWKPPANDGGAPIDAYIVEKKDKFGDMVBCARYD 3490  
QY 247 G-----F-----A-----A-----RI 251  
Db 3491 GKTATADNLTPGETYQFRVKAVNKAQSGKPSDPTGNVAVKPRMAPKLNLAGLLDLRI 3550  
QY 252 -----W-TKD-G-----NL-A-S-T-L-262  
Db 3551 KAGTPIKLDIAFEGERPAPVAKWKANDATIDTARADVTNTPTSSAIIHIFSAVRGDTGVYK 3610  
QY 263 -----G-Q-----HK-G-----P-----I-269  
Db 3611 IIVENHGHKTAQCNTVLDVFPETPEGLIKIDEIHKEGCTLAWKPPDNGDGLVHIVE 3670  
QY 270 -----F-A-----L-----272  
Db 3671 KMDTSGTQWVGCTFPDCTAKVKNLVPGEYAFRVKAVNLQGESKLEAEPIIAKNQPD 3730  
QY 273 -----KWN-----KK-----G-----278  
Db 3731 VPDVDPKPEVDWDKDRIDIKWNPNTANNGGAPVTGYIVEKEKGSAINTEAGTKPTGTF 3790  
QY 279 --N-----F-----IL-S-S-AG-----285  
Db 3791 ADNLKPGYEFVRVIAVNAAGSPDSDPTDQITKARYLKPKILTASRKIKKAGTHNL 3850  
QY 286 -----VD-K-TT-----290  
Db 3851 EVDFIGAPDPTATWVDSGAALAPPELLVDKSSITTSIFPSSAKRADSGNYKLKVKNELG 3910  
QY 291 --I-I-----W-----D-----294  
Db 3911 EDEAIPEVIVQDRPSAPEGPLEVSDVTKDSVNLNKKPPKDDGAEISNYVVEKRDTKNT 3970  
QY 295 -----A-----H-----T-G-----E-AKQF-----304  
Db 3971 WVPVSAFVTGTSITVPKLTEGHEYEFVRVMAENTFGRSDSLNTPDEVLAKDPFGTGPGR 4030  
QY 305 -----P-----F-----306  
Db 4031 PEIVDTDNHDIDIKWDPDRNGGSPVDHYDIERKDAKTGRWIKVNTSPVQGTAFSDTRVQ 4090  
QY 307 --H-----SA-P-A-LD-----V-----314  
Db 4091 KGHTEYRVVAVNKAQGPQSDSSAAATAKPMHEAPKFDLDDGKEFRVKAGEPLVITIP 4150  
QY 315 -----D-W-----Q-----S-----318  
Db 4151 FTASQPDIDISWTEGKPLAGVETTSQTLVIPSTERSDGPVKIKAVNPYGEAEANIK 4210  
QY 319 -----N-----N-----320  
Db 4211 ITVIDKEGAPENITYPAVSRHTCTLNWDAPKDDGAEIAGYKIEYQEVGSQIWDKVPGLI 4270  
QY 321 --T-----F-----A-S-C-----ST-----327



Db	4271	SGTAYTVRGLEHQVYFRIRAEANVGLSDYCOGVPVWIKDPDPGAPSTBITGYDTN	4330	Db	5348	DMDHPTIRIVFDDHILITIPALSVREAGRYEYTVNSDGEATTGFWLNVTLGLPEAQP	5407
QY	328	-----D-----M-----CT-----	331	QY	421	-----D-S-----TV-----	424
Db	4331	QVSLAWNPPRDDGSPILGVYVERFEKRGGDWAPVMPVMVKGTECIVPGLHENETYQFR	4390	Db	5408	LHISNIGPSTATLWSRPPVTDGSKITSYVVEKRDLSKDEWTVTSNVKDMNYIVTGLFE	5467
QY	332	-----H-----V-C-----	334	QY	425	-----	424
Db	4391	VRAVNAAGHGPESNGSEPVTCRPVVEKPGADAPRVGKITKNSAELTNWPLRDGGAPID	4450	Db	5468	NHEYEFVSAQNEENGIGAPLVSEHPHIIARLPDPPTSPLEIVQVGDVYVTLWSQRP	5527
QY	335	-----KLQ-----	337	QY	425	-----RL-----W-----	427
Db	4451	GYIVEKKLGDNDWTRCNDKPRDTAPEVKNLGEKBEYEPVAVNSAGEGEPKPSDLV	4510	Db	5528	DGGRLRGYIVEKQEBEHDEWFCNQPNPNPNVFNLDGRKYRYRFAVNDAGLSDL	5587
QY	338	-----Q-----D-----R-----PI-----	342	QY	428	-----D-----V-----D-----R-----	431
Db	4511	LIEQPGRPIDINNLKIDITVRAGETIQIRIPVAGGNPKPIIDLFNGNSPIFENERTVVD	4570	Db	5588	AELDQTLFOASGSGEGPKIVSPLSDLNEEVGRCVTFECEISGSPREYRWFKCKELVDT	5647
QY	343	-----K-T-----F-----Q-----	346	QY	432	-----G-----I-----C-----I-----H-----	436
Db	4571	VNPGIEVITTTGSKRSDAGPYKISATNKYKDKCLNVFLDAPGKPTGPIRATDIQADA	4630	Db	5648	SKYTLINKGDKQVLIINDLTSDDADEVTCRATNSGSTRSTRANLRITKPRVFIPPKYHG	5707
QY	347	-----GH-----T-----N-E-V-----N	353	QY	437	-----TL-----	438
Db	4631	MTLSWRPPKONGDAITNYVVEKRTPGGDWTVGVGPTTLRVRLNDANTPYEFRVRAEN	4690	Db	5708	GYEAKGETIELKIPYKAYPQGEARWTKDGEKIENNSKFSITTTDDKFATLRISNASREDY	5767
QY	354	-----AI-K-----W-----D-----P-----	359	QY	439	-----	438
Db	4691	QYGVGPLETDDAIVAKNPDPFTPGAPQGEAVETSEEAITLQWTRPTSDGGAPIQGYVIE	4750	Db	5768	GEYRVVSVNSGSDSGTVNTVADVPPEPRFPPIENILDEAVILSWKPPALDGGSLVTNY	5827
QY	360	-----T-----GN-----AS-----C-----	362	QY	439	-----T-----KH-Q-----E-----PV-----	445
Db	4751	KREVGSTEWTKAAGNILDTHKRVTLGTPKKTVEFRVAANAAGQGEYSVNSVPITADNA	4810	Db	5828	TIEKREMGSMSPCAKSRVYTTTISGLRAGKQYEFRIIAENKHGQSKPCPTAPVLIPG	5887
QY	363	-----L-----	363	QY	446	-----Y-----S-----V-----	448
Db	4811	PTRPKINMGLTRDILAYAGERAKILVFAASPAKVPKTFSGENKISPTDPRVKVEYSDF	4870	Db	5888	DERKRERYDVDEQGIKVRGKGTVSSNYDNYVDFIWKQYVQVPEIKHDHVLHDHYDHEE	5947
QY	364	-----L-----AS-----C-----	367	QY	449	-----AF-----S-----P-----	452
Db	4871	LATLTIEKSELTDGLYFVELENSQSDSASIRLKVVDPKASPQHIRVEDIAPDCCTLYW	4930	Db	5948	LGTGAGVHVHRTERATGNNAFAKFMVTPHESDKETVRKEIQTMSVLRHPTLNLHDAPE	6007
QY	368	-----SD-----DM-----I-----W-----SMK-----Q-----	372	QY	453	-----D-----G-----R-----Y-----	456
Db	4931	MPSSDGGSPITWYIYEKLDLRHSDGKWEKVSFVRNLNTVGLIKDNRYPFRVRAEQ	4990	Db	6008	DDNEMWMIYEFNMSGGELFEKVADEHNKMSDEAVEYMRQVCKGLCHMHENNYVHLDLKE	6067
QY	373	-----L-----K-----I-----W-----SMK-----Q-----	380	QY	457	-----L-----A-----	458
Db	4991	YGVSEPELADVVAKYQFEPVNPQPEAPTVRDKDSTWAELEWDPDRDGS-KIIGYQVQY	5049	Db	6068	NIMFTTKRSNELKLIIDFGLTAHLDPKQSVKVTGTGTAFAAPEVAEGKPVGYTDMMSVG	6127
QY	381	-----N-----CVHD-----L-Q-----A-----	389	QY	459	-----SG-S-F-----D-----K-C-----	465
Db	5050	RDTSGRWINAKMDLSEQ-HARVTLRQNGEPEFRRIIKNAAAGFSKPPSPERCQLKER	5108	Db	6128	LSYLLSGUSPFGENDDETNRNVKSCDNMDDSAFSGISEDGKDFIRKULLADPNTRMT	6187
QY	390	-----H-----N-----K-----EI-Y-TIKW-----S-----	400	QY	466	-----VH-----I-----W-----N-----	470
Db	5109	FGPPGPIHVAGSIGSNHCTIITWMAPLEDGGSKIITGVNVEIREYESTL-WTVASDYNVR	5167	Db	6188	IHQALEHPLWTCNAPGRDSQIPSSRYTKIROSITKYDAWPEPLPLGLGRISNYSILRKH	6247
QY	401	-----P-T-----GP-----G-----T-----	406	QY	471	-----TQ-----T-----G-A-L-----V-H-----	478
Db	5168	EPEFTVDKLEFNDYEFVRVAINAAGKIPSLPSGPIKIQESGSRPQIIVKPEDTAQPY	5227	Db	6248	RPOEYSIRDAFWDRSEAQRPFIVKPYGTEVGEQSANFYCRVIASSPPVVTMHKDDRELK	6307
QY	407	-----N-----P-----	409	QY	479	-----S-Y-R-----G-T-----G-----G-----I-----FE-----	488
Db	5228	NRAVFTCEAVGPEPTARLNRGRELPESSRYRFEASDGVYKFTIKEVMDIDAGSYTVE	5287	Db	6308	QSVKMYKNGYNDYGLTINRVKDDKGEYTVRAKNSYGTKEIIVFLNTRHSHSPLKEPL	6367
QY	410	-----ANL-----	413	QY	489	-----V-----C-----	491
Db	5288	VSNPYGSDTANLVQAPPVIEKDVNTILPSGLVRLKIYPSGTAPFRHSLVLRBEI	5347	Db	6368	EPKKKAPSPRVEEFKERSAPFTFHLNRLLIQNHQCKLTCSLOGNPNTIEWKDGH	6427
QY	414	-----L-----A-S-----A-----S-----F-----	420	QY	492	-----NA-----AG-----D-----K-----	497
Db				Db	6428	PVDEDRVQVFRSGVCSLBIIFNARVDDAGTYTATNDLGDVDVSECVLTVQTKGSPIPR	6487

QY 498 V-----G-----A-----SA-----502  
Db 6488 VSSFRPRAYDTLSTGTDVSHSYADMRRLSRDVS PDVRSAADLTKITNLPST 6547  
QY 503 ---SD---GS-----V-----507  
Db 6548 AQLSDSETVGGSAEFAAVSGQPEPLIEWLHNGERISEDSRFRASYVAGKATLRISA 6607  
QY 508 -----C-----VL---D-----LR--K 514  
Db 6608 KKSDEGYLCRASNSAGQEQTRATLT VKGDQPLNGHAGQAVESELRVTK 6657

RESULT 12  
T27935  
hypothetical protein ZK617.1b - Caenorhabditis elegans  
C;Species: Caenorhabditis elegans  
C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C;Accession: T27935; T28031  
R;White, S.  
submitted to the EMBL Data Library, May 1996  
A;Reference number: Z20442  
A;Accession: T27935  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-7160 <WIL>  
A;Cross-references: UNIPROT:Q23551; EMBL:Z73897; PIDN:CAA98065.1; GSPDB:GN00022; CESP:ZK617  
A;Experimental source: clone ZK617  
R;Harris, B.  
submitted to the EMBL Data Library, May 1996  
A;Reference number: Z20458  
A;Accession: T28031  
A;Status: preliminary; translated from GB/EMBL/DBJ  
A;Molecule type: DNA  
A;Residues: 1-7160 <WIL>  
A;Cross-references: EMBL:Z73899; PIDN:CAA98082.1; GSPDB:GN00022; CESP:ZK617.1b  
A;Experimental source: clone ZK829  
C;Genetics:  
A;Gene: CESP:ZK617.1b  
A;Map position: 4  
A;Introns: 10/3; 61/3; 135/2; 168/3; 256/2; 379/3; 405/2; 463/1; 508/3; 542/3; 574/3; 59  
3067/1; 3141/3; 3269/1; 6473/3; 7012/3; 7097/1; 7129/3  
C;Superfamily: twitchin; fibronectin type III repeat homology; immunoglobulin homology;

Query Match 70.5%; Score 2585; DB 2; Length 7160;  
Best Local Similarity 8.0%; Pred. No. 5.9e-52;  
Matches 470; Conservative 31; Mismatches 9; Indels 5360; Gaps 386;

QY 1 M-----S-----I-----S-S-----D-----6  
Db 1113 MTEARRQSLPFGKKVEKWDIPLPKTVQQQVDKICEWKCTYSRPNAKIRWKDKRKEIFS 1172  
QY 7 -----EV-----N-----FL-----11  
Db 1173 GGLKXKIVIEKNVCTLIINNEVDVTGKYTCENGVPHTAQLTVLEPPMKYSFLNPLNT 1232  
QY 12 ---VTR---Y-----LOE-----S-G-----20  
Db 1233 QEIVRTKQAVLTCKVNTPRAPLVWYRGSKAIQEGDPRFIIIEKDAVGRCTLTIKEVEBDQ 1292  
QY 21 -----FS-----HSAP-----T-----27  
Db 1293 AEWTARITQDVFSKVQVTVVERPRT-FVVPKMSQKNESDLATLETLVNDKDAEVVWHD 1351  
QY 28 -----F-----G--IESH-----33  
Db 1352 GKRIDIDGVKVESNRKRLLINGARIEDHGEYKCTTKDRDTMAQLIVDAKNKFIVAL 1411  
QY 34 -----IS-----QS-N-----I--N 40  
Db 1412 KDEVIEKDDVTLMQTKDTKTPGIWFRNGKQISSMPGKPFETOSRNGHTHTLKGIEMN 1471

QY 41 -----G-----A-----L-----VP-----P 46  
Db 1472 EADVYIEDQALRGSNCVTVLEAEKRPILNWKPKIEAKAGPCPVVVKFQIKGTRRGDP 1531  
QY 47 -AA-L---IS-----II-----Q-----54  
Db 1532 KQAILKNGKPIDEEMRKLVEVITKDDVAEIVFKNPQLADTGKWALELNSAGTALAPPEL 1591  
QY 55 -----KG-LQ-----VV-----E-----A-----62  
Db 1592 FVKDKPKPPKPLETKNTVTAEGLDLVWGTPDDEGAPVKAYIIEMOEGSRGNWAKVGETK 1651  
QY 63 -----E--V-----SI-----N-----67  
Db 1652 GTDFVKDLKEHGEYKFRVKALNECGLSDPLTGESVLAKNPYGVGPKQNMDAIDVDKDH 1711  
QY 68 -----EDG-----T-----L-----72  
Db 1712 CTLAWEPPEDGGAPITGYIIRERREKSKDHWGQTKPDCCCELTDKKVVEDKEYLYRVK 1771  
QY 73 -----F-----D-----74  
Db 1772 AVNKAGPGDCHGKPIKMKAKKASPEFTGGGIGKDLRLKVGETIKYDVPISGEPLPECLW 1831  
QY 75 ---GRP---IE-----S-----L-----82  
Db 1832 VVNGKPLKAVGRVMSRSGKHKIMKENAVRADSGKFTITLKNSSGSCDSTATVTVGPR 1891  
QY 83 ---L-I---D-AVM---PD-----VV-Q-----T 94  
Db 1892 TPKGGLDIADVCADGATLSMNPDDGGDPLTGYIVEAQDMDNKGYIEVGKVDNNTTT 1951  
QY 95 -----R-----Q-Q-----97  
Db 1952 LKVNGLRNKGYFRVKAVNNEGESEPLESADQYTIKDPWDEFGKPRPEITDPADRID 2011  
QY 98 -----A---Y---RD-----101  
Db 2012 IAWEPHKGGAPIEYIIVEVRDPTDKWKVKVRVDPDTNASISGLKEGKEYQFRVAVANK 2071  
QY 102 ---K-LA-----104  
Db 2072 AGPGQSEPEKQAKRPFIPAWLKHDNLKSI TVKAGATVRWEVKITGGSEPIPEVKWFGN 2131  
QY 105 QQ-----H-----A-----A-----110  
Db 2132 QQLENGIQLTIDTRKNEHTILCIPSAMRSDVGEYRLTVKNSHGADSEKANTVLDRPSKP 2191  
QY 111 -----A-----A-----A-----113  
Db 2192 NGPLEVSDVPEDNLNLWSKPPDDGGEPTEYVEVEKLTATGRWVPCAKVKOTKAHIDGL 2251  
QY 114 -----A-----AA-A---T---N-----119  
Db 2252 KKGQTYQFRVKAVKESGASDALSTDKDTAKNPYDEPGKTGTPDVPVDWADRVLSLEWPP 2311  
QY 120 -----Q-----Q-G--S-----A-----K-N---G 127  
Db 2312 KSDGGAPITQYVTEKKKGKGRDQWQEGKVSQDQTNAEILGLKEGEBEYQFRVKAVNKAGPG 2371  
QY 128 E-----N-----T---AN-----GE-----E---N-----136  
Db 2372 EASDPSRKVAKPNLKPWIDREAMKTIITIKVGNDFEFDVPRVGEPPPKKEWIFNEKFPVD 2431  
QY 137 -----GA--H---TI-A-N-----143  
Db 2432 DQKIRIESEDKTRFVLRGATRKHAGLYTLTATNAGSDKHSVEVIVLGPSSPLGLEV 2491  
QY 144 -N-----H-----TDM-----ME- 150  
Db 2492 SNVYEDRADLEWKVPEDGGAPIDHVEIEKMDLATGRWVPCGRSETTKTTVPNLQPGHEY 2551  
QY 151 ---V---D-----G-----D---154

Db 2552 KFRVAVNKEGESDPLTTNTAILAKNPYEVGVKPELVWDKOHVDLAWNAPDDGGAP 2611  
QY 155 VE--I--P--S--NKA--162  
Db 2612 IEAFVIEKKDKXGRWEALVVPDQKATVPNLKEGEYQFRISARKAGTGDPSDSR 2671  
QY 163 VW--L--R--G--H--ES--170  
Db 2672 VVAKPRNLAPRIHREDLSDTTVKVGATLKPIVHIDGEPADVTWSPNGKIGESKAQIEN 2731  
QY 171 -----F--I--C--EV--172  
Db 2732 EPYISRALPKALRKOSGKYTITATNINGTDSVTINIKVKSKPKGPIEVTVDFEDRA 2791  
QY 173 -----F--I--C--EV--175  
Db 2792 TLDWKPPEDDGGPIEFYIEIEKANTKDWIWPGRSGDTHFTVDSLNGDHYKFRVAVN 2851  
QY 176 -----A--176  
Db 2852 SEGSPDLETTDILAKNPFDRPGRPEPTDWDSDHVDLKWDPPLSDCGAPIEEYQIE 2911  
QY 177 -----W--177  
Db 2912 KRTYGRWEPAITVPGGQTTATVPDLTPNEEYFRVAVNKGSPDPSDASKAVIAKPRN 2971  
QY 178 -----N--PVS--D--L--183  
Db 2972 LKPHIDRALKNUTIKAGQISFDVPVSGEPAPVTVTHWPDNREIRNGRVLNDPEYOS 3031  
QY 184 -LV--SG--S--GD--190  
Db 3032 KLVVKQMERGDSGTFITKANWANGEDEATVKINVIDKPTSPNGPLDSDVHGHDVTLNWR 3091  
QY 191 -----S--TA--R--194  
Db 3092 APDDGGPIENYVIEKYDTASGRVPAKAVAGDKTTAVDGLIPGHEYKFRVAAVNAEG 3151  
QY 195 -----I--195  
Db 3152 ESDPLETFTLAKDPFKGKNAPEITDWDKHDVLEWKPPANDGGAPIEEYVEMKD 3211  
QY 196 -----W--NL--S--E--200  
Db 3212 EFSPFNDVAHVAGQTNATVGNLKEGSKYEFRIAKNAGLGDPSDSASAVAKARNVP 3271  
QY 201 -----NS--TSG--PT--Q--L--V 210  
Db 3272 VIDRNSIQEIKVKGQDFSLNIPVSGEPTTITWTPEGTPEVSDRMLNEDGKTFHV 3331  
QY 211 -----LR--H--C--214  
Db 3332 KRALRSDTGTYYIIKAENENGTDFAEVKVTVLDPSPFRGLDVTNIVKDCDLAWKEPED 3391  
QY 215 -----I--215  
Db 3392 DGAIEISHYVIEKQDAATGEWTAGESKOTNFHVDDLTOGHEYKFRVAVNRHGDSDPLE 3451  
QY 216 -RE--217  
Db 3452 AREALIAKDPFDRADKPGTPEIVWDKHDADLKWTTPADDGAPIEGYLVEMTPTSGDW 3511  
QY 218 -----G--218  
Db 3512 PAVTVGAGELTATVDGLKPGQTYQFRVAKNKGESTPSPSRMTVMKPRHLAPKINRDM 3571  
QY 219 -----GO--DV--P--223  
Db 3572 FVAQRVAKAGTQLNFDVNVGEPEAPKIEWFLNGSLSSGNGTHIDNNTKLTTKSTAR 3631  
QY 224 -----S--N--KDV--S--LD--232

Db 3632 ADSGKYIVATNESGKDEHEVDVNIILDIPCAPEGPLRHKIDITKESVVLKWDEPLDDGGSP 3691  
QY 233 -----W--N--S--GT--L--239  
Db 3692 ITNVYVEKQEDGRWVPCGETSDTSUKVNLSEGHYKFRVAVNRQGTSAPLTSDHAI 3751  
QY 240 -----L--A--T--G--SY--D 246  
Db 3752 AKNPPDEPDADPTDVTVDWDKOHVDLEWKPANDGGAPIDAYIVEKKDKFGDWVECARVD 3811  
QY 247 G-----F-----A-----RI 251  
Db 3812 GKTTKATADNLTPGETYQFRVAVNKGKPGKPSDPTGNVAVKPRMAPKLNLAGLIDLRI 3871  
QY 252 -----W--TKD--G--NL--A--S--T--L--262  
Db 3872 KAGTPIKLDIAPEGEPAVPAKWKANDATIDTGARADVTTPTSSAIIHIFSAVNGDGTGVYK 3931  
QY 263 -----G--Q-----HK--G-----P-----I--269  
Db 3932 IIVENEHGKDTAQCNVTVDLPCTPEGLKIDIEIHKEGCTLNWKPPTDNGGTDVLHVIVE 3991  
QY 270 -----F--A-----L-----272  
Db 3992 KMDTSRGTVQEVGTPDCTAKVNLVPGKEYAFRVAKNLQGESKPLEABEPIIAKNQFD 4051  
QY 273 -----KWN-----KK-----G-----278  
Db 4052 VPDVDKPEVTDWDRIDIKNPTANNGAPVTGVIIVEKKEGSAIWEAGTGTPTFS 4111  
QY 279 --N--F-----IL--S--AG-----285  
Db 4112 ADNLKPGVEYFRVIAVNAAGPSDPTDPOITKARYLKPILKTASRKIKIKAGFTHNL 4171  
QY 286 -----VD--K--TT-----290  
Db 4172 EVDFIGADPTATVTVGDSGAALAPELLVDKSTSTSIFFPSAKRADSGNYKLVKNELG 4231  
QY 291 -----I--I--W-----D-----294  
Db 4232 EDEAFEVIVQDRPSAPEGLEVDVKDSCVLNWKPPKDDGAEISNYVVEKRDYKNT 4291  
QY 295 -----A-----H-----T--G-----E--AKQF-----304  
Db 4292 WVPVSFAVTGTSITVFKLKEGHEYFRVMAENTFGRSDSLNTDEPLAKDPFGTGPCKGR 4351  
QY 305 -----P-----F-----306  
Db 4352 PEIVDTDNHDIDIKWDPPRDNGSPVDHYDIERKDAKTGRWIKVNTSPVQGTAFSDTRVQ 4411  
QY 307 --H-----SA-----P--A-----LD-----V-----314  
Db 4412 KGHTEYRVAVNKGAPGQPSDSSAAATAPKMPHEAPKFDLDLKGKFRVAKGEPVITIP 4471  
QY 315 -----D--W-----Q-----S-----318  
Db 4472 FTASQPDLSWKEGKPLAGVETDTSQTLVPISTRSDSGPVKIKAVNPYGEAEANIK 4531  
QY 319 -----N-----N-----320  
Db 4532 ITVIDKPGAPENITYPAVSRHTCTLNWDAPKDDGAEIAGYKIEYQEVGSIQIWKVPGLI 4591  
QY 321 -----T-----F--A--S--C--ST-----327  
Db 4592 SGTATVIRGLEHQYFRIRAEAVGLSDYCOGVPVWIKDPPDPGAPSTPEITGYDTN 4651  
QY 328 -----D-----M-----CI-----331  
Db 4652 QVSLAWNPRDDGSPILGYVVERFEKGGDWAPVMPMPVKGTGECIVPGLHENETYQFR 4711  
QY 332 -----H-----V--C-----334  
Db 4712 VRVNAAGHGPESNGSEPVTCRPPYVEKPGADAPRVGKITKNSAELTNRPLRDGGAPID 4771

QY 335 -----KLG----- 337  
Db 4772 GYIVKKGKLDNDWTRCNKPDRTAFVKNLGEKEBEYFRVIAVNSAGEBEPKPSDLV 4831  
QY 338 -----Q-----D-----R-----PI----- 342  
Db 4832 LIBEQGRPIFDINNLKDITVRAGETIQIRIPVAGGNPKPIIDLFGNNGPIFENERTVVD 4891  
QY 343 -----K-T-----F-----Q----- 346  
Db 4892 VNFGEIVITTTGSKRSAGPYKISATNKYKDKTCKLVNFDAPGKPTGPIRATDIQADA 4951  
QY 347 -----GH-----T-----N-B-V-N 353  
Db 4952 MTLWRPPKNGGDAITNVVVEKRTGGDWTVGHPVGTTLVRNLDANTPYEFRVAEN 5011  
QY 354 -----AI-K-----W-----D-B- 359  
Db 5012 QYGVGBLETDDAIVAKNPPDTPGAPGQPEAVETSEAITLQWTRPTSDGAPIQGYVIE 5071  
QY 360 -----T-----GN----- 362  
Db 5072 KREVGSTEWTKAAGNILDTHRVGTGLTPKTYEFRVAAYNAAGQGEYSVNSVPIADNA 5131  
QY 363 -----L----- 363  
Db 5132 PTRPKINMGMLTRDILAYAGERAKILVPPFAASPAPKVTFSKGENKISPTDPRVKVYSDF 5191  
QY 364 -----L-----AS-----C----- 367  
Db 5192 LATITKESLTDGLYFVLENSQGSASIRLKVVDKSPQHIRVEDIAPDCCLYW 5251  
QY 368 -----SD-----DM----- 372  
Db 5252 MPSSDGGSPITNIVIVEKDLRHSQKWKVSVFVRLNVTVGGLIKDNRYFRVRAETQ 5311  
QY 373 -----L-----K-----I-----W-----SMK-----Q----- 380  
Db 5312 YGVSEPCELADVVAKYQFEVNPQPEAPTRVDKSTWAELEWPPRDGGS-KIIGYQVQY 5370  
QY 381 -----D-----N-----CVHD-----L-Q-----A----- 389  
Db 5371 RDTSSGRWINAKMDLSQC-HARVTGLRQNGEPEFRIIAKNAAGFKSPSPSERCQLKSR 5429  
QY 390 -----H-----N-----K-----EI-Y-TIKW-----S----- 400  
Db 5430 FGPPGPIHVGAKSIGRNHCTITWMAPLEDGGSKIYGNVEIREYGSTL-WTVASDYNVR 5488  
QY 401 -----P-T-----GP-----Q-----T----- 406  
Db 5489 EPEFTVDKREFNDYEFRRVAINAAGKIPSLPSGPIKIOESGGRPQIVVKPEDTAQPY 5548  
QY 407 -----N-----N-----P----- 409  
Db 5549 NRRAVTCEAVGREPTARMLNRGRELPESSRYRFEASDGVYKTIKEVWDIDAGEYVE 5608  
QY 410 -----N-----ANL----- 413  
Db 5609 VSNPYGSDTATANLVQAPPVIEKDVNTILPSGLVRLKIYFGTAPFRHSLVLRNEEI 5668  
QY 414 -----M-----L-----A-S-----A-----S-----F----- 420  
Db 5669 DMDHPTIRIVEFDHILITIPALSVREAGRYEYTVSNDSGEATGFWLNVGTGULPEAPQGP 5728  
QY 421 -----D-S-----TV----- 424  
Db 5729 LHSNIGPSTATLSWRPPVTDGSKITSYVVEKEDLSKDEWTVTSNVKDMNYIVTGLPE 5788  
QY 425 ----- 424  
Db 5789 NHEYEFVSAQNGENGIGAPLVSEHPHIIARLPDPPTSPNLNLEIVQVGDDYVTLWSORPLS 5848

QY 425 -----EL-----W----- 427  
Db 5849 DOGGLRLGYIVKEQBEEDHEWFRQNFQNPNNVFNLI DGRKYRYRFAVNDAGLSDL 5908  
QY 428 -----D-----V-----D-----R----- 431  
Db 5909 AELDTQLFOAGSGGPKIVSPLSLDNEBVGCVTFECEISGSPREYRWFKCKELVDT 5968  
QY 432 -----G-----I-----C-----I-----H- 436  
Db 5969 SKYTLINKGDKQVLIINDLTSDDADEYTCRATNSGSTRANLRITKTRPVFIPPKYHG. 6028  
QY 437 -----TL----- 438  
Db 6029 GYBAQGETIELKIPYKAYPQGEARWKDGEKIENNSKPSITTTDDKFAFLRISNASREDY 6088  
QY 439 ----- 438  
Db 6089 GEYRVVVENSGSDGTNVNVTADVPEPPRPPIENILDEAVILSMKPPALDGGSLVTNY 6148  
QY 439 -----T-----KH-Q-----E-PV----- 445  
Db 6149 TIEKREAMGGSWSPCAKSRYTYTTIEGLRAGQYEFRIIAENKHGOSKCPCEPTAPVLIPG 6208  
QY 446 -----Y-----S-----V----- 448  
Db 6209 DERKRRGYDVDEQKIVRGKGTSSNYDNYVDIWKQYYPQVPEIKHDHLDHYDIHEE 6268  
QY 449 -----AF-----S-----P----- 452  
Db 6269 LGTGAPGVHVRTERATGNFAAKFVMTPHESDKETVRKEIQTMSVLRHPTLLVNLDAPE 6328  
QY 453 -----D-----G-----R-----Y----- 456  
Db 6329 DDNEMVMIYEFMSGGELFEKVADEHNKMSDEAVEYMRQVCKGLCHMHNYYVHLDLKEE 6388  
QY 457 -----L-----A----- 458  
Db 6389 NIMFTTKRSELKLDLDFGLTAHLDPKQSVKVTGTGTAFAAPEVAEGKPVGYTDMWSGV 6448  
QY 459 -----SG-S-F-----D-----K-C----- 465  
Db 6449 LSYLLSGLSPFGGNDDETLRNVKSCDWNMDSAFSGISEDGKDFIRKLLADPNTRMT 6508  
QY 466 -----VH-----I-----W-----N----- 470  
Db 6509 IHOALSHPLWTPGNAPGRDSQIPSSRYTKIRDSIKTKYDAWPEPLPLGRISNYSRLKH 6568  
QY 471 -----TQ-----T-----G-A-L-----V-H----- 478  
Db 6569 RPOEYSIRDAFWDRSEAQRPFIVKPYGTEVGEQSANFYCRVIASSPPVVTWHKDDRELK 6628  
QY 479 -----S-Y-R-----G-T-----G-----I-----FE----- 488  
Db 6629 QSVKMKRYNGNDYGLTINRVKGGDKGEYTVRAKNSYGTKEEIVFLNTRHSEPLKFEPL 6688  
QY 489 -----V-----C-----W----- 491  
Db 6689 EPMKAPSPRVEEFKERRSAPFTTHLRNLIQKNHQCKLTCSLQGNPNPTIEWMKDGH 6748  
QY 492 -----NA-----AG-----D-----K----- 497  
Db 6749 PVDEDRVQVSFRSGVCSLEIFNARVDDAGTYTATNDLGDVDSVCEVLTVTQKGGEPIPR 6808  
QY 498 -----V-----G-----A-----SA----- 502  
Db 6809 VSSFRPRAYDILTSTGTVDERSHYADMRRLSRIRDVSPDVSRAADDLTKITINELPSFT 6868  
QY 503 -----SD-----GS-----V----- 507  
Db 6869 AQLSDSETEVGSAGAEFSAAVSGQPEBLIEWHNGERISEDSRFRASYAGKATLRISDA 6928  
QY 508 -----C-----VL-----D-----LR-K 514



Db	1852	KGTRVKGKTLIANDGTPLGVGPPVGLAVDPARGKLYWSDHGTDSGVPKAKIASANMDGTSI	1911
Qy	200	E-----N-----S-----TS-----G-----P-----	206
Db	1912	KILFTGNLQHLVVLTDIOEQKLYWATSRGVIERNVNDGTERMIHLVHLAHPGLWVYG	1971
Qy	207	-----T-----Q-----VLR-----H-----C-----	214
Db	1972	SFLYSDQYEVIERVDKSGNNKVVLRDNPVYLRGLRVHRRNAADSSNGCSNPNACQ	2031
Qy	215	I-----R-----E-----G-----	218
Db	2032	QICLPVPGMFWSCACAGFKLSPDGRSCSPYNSFMVVMPLPAVYRGSLELSHDSEAMVPV	2091
Qy	219	-GO-----DV-----P-----SN-----K-----	226
Db	2092	AGQGRNVLHADVDVANGFIWCDPSSSVSSNGIRRIKPDGSNFTNVVTVYGANGIRGV	2151
Qy	227	-D-----V-----TS-----L-----D-----W-----	233
Db	2152	ALDWAAGNLYFTNAFVETLIEVLRINTTYRRVLLKVSVDMPRHIIIVDPKHYLFWADYG	2211
Qy	234	-----N-----SEG-----T-----LLA-----TGSY-----DG-----	247
Db	2212	QPKIERSFLDCTNRTVLVSEGIPTPGL-AMHDGTG-YIYWDDSLDLIARIHLDGSES	2269
Qy	248	-----F-----AR-----TWKD-----GN-----	257
Db	2270	QVRYGSRYPYPTGITVFGESIIV-DRNLKVKFQASKQPGNTDPPVIRDKINLRDVT	2328
Qy	258	-----L-----A-----S-----TLG-----O-----	264
Db	2329	IFDEHAQPLSPAELNNPCLQSGGCSHFCAFPALPELPTPRCGCAFGTLGNDGSKATSQE	2388
Qy	265	-----H-----K-----	266
Db	2389	DFLIYSLNLSLHFDPRDHSUPFQVISVAGTALDLYDRNRNRIFFTKLNSLRQIS	2448
Qy	267	-----G-----P-----I-----F-----	270
Db	2449	YVSLYSGSSPTVLLSINIGVTGDIADFWINRRIYSDFSQNTINMAEDGNSRAVIARVS	2508
Qy	271	-----A-----LK-----W-----N-----K-----K-----GNF-----IL-----S-----AG-----	285
Db	2509	KPRAIVLDPCRGYMTDWTGNKIERATLGGNFRVPIVNTSLVWPNGLALDLETDLLYW	2568
Qy	286	VD-----K-----T-----TI-----I-----W-----D-----	294
Db	2569	ADASLQIERSTLTGTNRVVVSTAFHSFGLTVYVQYIYWTDLTKIYRANKYDGSGLV	2628
Qy	295	A-----H-----TG-----E-----A-----	300
Db	2629	AMTTRLPQSGISTVVKTORQCSNCPDQFNGGCSHICAPGNARGCQCPHEGNWYLAN	2688
Qy	301	-----K-----O-----OF-----P-----PH-----S-----A-----	309
Db	2689	DNKYVVDVTGRCNQLOFTCLNGHCINQDWKCDNDNDGCGSDELPTVCAFHCTRSTFT	2748
Qy	310	-----P-----A-----L-----	312
Db	2749	CGNRCVPYHYRCDYNDGDSDEAGCLPNCNMTTEFTCSNGRCIPLSVYVCGINNCH	2808
Qy	313	-----D-----V-----D-----	315
Db	2809	DNDTSDEKNCPPTCPDFTKQCTTNICVPRAFLCDGNDGCGSDENPIYCASHTCRSN	2868
Qy	316	-----W-----O-----S-----	318
Db	2869	BFQCLSPORCIPSWFCDEADGADGDEPDTCGHSUNTCRASQFQCDNGRCISGNWVCD	2928
Qy	319	-----N-----T-----F-----	322
Db	2929	GDNDGDMDEQDRHHCELQNCSTQFTCVNSRPPNRRCIPOYWVCDGDADCDALDELQ	2988
Qy	323	-----A-----SC-----	325
Db	2989	NCTMTCSAGBFSCANGRCVROSFCDRRNDGCDYSDERGCSYPPCHANQFTCONGRCP	3048
Qy	326	-----S-----T-----D-----MCI-----HV-----C-----	334
Db	3049	REFVCDNDGDSGEQEHLCHTPEPTCLPHQFRCDNGHCEMGRCVCHNVDDCSDNSDE	3108
Qy	335	K-----L-----G-----Q-----	338
Db	3109	KGGINECLDSSISRCDHNTDITITSFYCSCPLGYKLMDSKSCVDIDECKESPQLCSOK	3168
Qy	339	-----D-----R-----P-----I-----	342
Db	3169	CENVVGSYICKAPGYIREPDGSKCRQNSIBPYLIFSNRYIRNLTDTGSSYLILOGL	3228
Qy	343	-----KT-----	344
Db	3229	GNVVALDFRVEKRLYWDIAEKQIIERMFLNKTNRETIINHRLRAESLAVDWVSRKLYW	3288
Qy	345	-----F-----Q-----GH-----	348
Db	3289	LDAILDCLFVSDLEGRHRRKMIQAHCVDANNTFCFEHPRGIVLHPQRGHVYVADWGVHAYI	3348
Qy	349	-----TNE-----V-----NAI-----KW-----D-----	358
Db	3349	GRIMDGTNKSIIITKIEWPNAITIDYNDLLYADAHGLGYIEFSDLEGHHRHTVYDGS	3408
Qy	359	-----P-----T-----GN-----L-----LAS	366
Db	3409	LHPFALTIFEDTVFTWTDWNTRTVEKNGYDGSGRVLTNTTHKPPDIHVYHPYQPIMS	3468
Qy	367	-----CS-----DD-----M-----T-----L-----K-----	374
Db	3469	NFCGTNNGCSHLCLIKAGRGFTACPDPTQVQLRDLRCLCMPCMSQFCLGNNKCI	3528
Qy	375	-TW-----S-----MK-----Q-----D-----NC-----	383
Db	3529	PIWKCQDQKDCSDGDEPDLCPHRCPLGQPCQDGNCTSPQALCNARQDCADGSEDR	3588
Qy	384	V-----H-----D-----L-----Q-----AH-----N-----	391
Db	3589	VLCHEHRCESNEWQCANCKICIPQSWQCDVNDCLDSDSDTSHCASRTCRPGQFKCNGR	3648
Qy	392	-----K-----E-----I-----Y-----T-----IK-----W-----	399
Db	3649	CIPQSWKCDVNDGCDYSDEPIDECTTAAAYCNDHTEFCKTNYRCIPQWAVCNGFDDCR	3708
Qy	400	-----S-----PTG-----P-----GT-----N-----N-----	408
Db	3709	DNSDEGCSVPCHPSGDFRCANHHCIPLRWKCDGTDGCDNSDEENCVPRECSSEFRC	3768
Qy	409	-----P-----N-----AN-----LM-----LA-----SA	418
Db	3769	ADQOCIPSRWVCDQDNDGSDERDCMKTCHEPHFQCTSGHCVKALACDGRADCLDA	3828
Qy	419	SFD-----S-----TVR-----LW-----D-----VDRG-----I-----C-----I-----	435
Db	3829	S-DESACPT-RFPNGYCPAAMPECKNHCIOFQWICDGENDCVD-GSDEEHLCLFNPIC	3885
Qy	436	-----HTL-----T-----K-----H-----QBPV-----Y-----S-----	447
Db	3886	ESQPRFCDNSRCVYGHQLCNGVDDCGDGSDEKEEHCRTKPTDTEYKCSNGNCISQ	3945
Qy	448	-----V-----A-----F-----S-----P-----	452
Db	3946	HYVCDNVDCGLDSETGCLGDNRTCAENICEQNTQLSGGGFCISCRPGFKPSTDKN	4005
Qy	453	-----D-----G-----VILAS-----GS-----P-----D-----K-----C-----	465
Db	4006	SCODINECEBFGICPOSCR--NSKGYECFCVDGFKSMSTHYGERCAADGSPFLLLPEN	4063

466 V-----HI-----WN-----T---Q-T-GA-----475  
4064 VRIRKNTSSEKFELEEHIOTIDYDNDPEHIGLSVVYVTVLAQSGOFGAKRAYIP 4123  
476 -----L-----V-H-----S-----YR-----481  
4124 NFESGNNPIREVDLGLKLMQPDGLAVDWGRHIYWSDAKSORIEVATLDGRYRWLIT 4183  
482 -----G-----T-----G-----G-----484  
4184 TQLDOPAAIAPNPKLGLMFTDQKQPKIESAMWNGEHRVLSNGLWPNGLSIDVND 4243  
485 -----G-----G-----IFE-----V-----C-W--NAAG 495  
4244 DRVWSDSKEDVIEAIVDGTDRRLIINEAMKPFSLDIPEDKLYWVAKEGEVWRQKFG 4303  
496 -----DKV-----LD-----GA-S-A-----SD-----504  
4304 KENKEVLVVPMLTVQVIFHQLRYNSQSVNPKQVCSHLCLLRPGGYSCACPGSDFVT 4363  
505 GS-V-C-----V-----LD-----511  
4364 GSTVQCDAASELPTVMPPPCRMHGNCYFDENELPKCKCSSLGSGYSGYCEVGLSRGIPPG 4423  
512 -----LTK 514  
4424 TTMAVLLTFVITVIGALVLVGLFHYRK 4451

RESULT 14  
T12117  
polyprotein - fava bean dsRNA replicon  
C:Species: Vicia faba (fava bean)  
C:Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
C:Accession: T12117  
J. Gen. Virol. 79, 2349-2358, 1998  
A:Title: Nucleotide sequence, genetic organization and expression strategy of the double  
A:Reference number: 217424; MUID: 98451319; PMID: 9780039  
A:Accession: T12117  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-5825 <PFE>  
A:Cross-references: UNIPROT:O82731; EMBL:AJ000929; NID:g3184155; PIDN:CA04392.1; PID:g3  
A:Experimental source: virion; cultivar 447  
C:Comment: This gene product may be cleaved into several proteins including helicase and  
C:Genetics:  
A:Genome: dsRNA replicon  
C:Superfamily: fava bean dsRNA replicon polyprotein

Query Match 70.1%; Score 2580.1; DB 2; Length 5825;  
Best Local Similarity 9.1%; Pred. No. 5.2e-52;  
Matches 442; Conservative 56; Mismatches 13; Indels 4349; Gaps 378;  
1 M-----S-----D-----E-----V-----N-- 9  
949 MATPESITQDAPVKESNTLLIDMNLMDPMWLITECLRRKITYATVTPMLELQNPD 1008  
10 -----F-----LV-----Y-----13  
1009 QTILDEGVQFRKVENQIVITAWSRPIRVTTENYMNWTKVDVAFIGENTYPIDTVREAN 1068  
14 -----R-----Y-----L-QE- 18  
1069 GISLRIDLAASSTFENSEVETVGSIEISVPTVYVYGGPKRKLMSGGEKKVKLHREL 1128  
19 -----SG-----FS-----HS-AP-----26  
1129 YRPMIRNVSGTSPKELVDYIGISNMKNINDMVVSFSDSDITRHTAAFYSKMIT 1188  
27 -----T-----F-----28

1189 RANSMOLTERVLESVNTARGFLASLLISOAETMLKITGLDEMTLTTVNRWMSKSEMVRL 1248  
29 -----G-----G-----I-----E--- 31  
1249 KGLMNLNSLDIDNLMCYSRPVRHLRAOETLELEPTGTCPHKTPDFVKECSMCCECG 1308  
32 -----S-----H-----33  
1309 VNPAIGAIGFSSCLPTDKHQCDHPCKHAHESEAGTRCSCCLPIAGEACPCCGVNRQIE 1368  
34 --I-----S-----Q-----SNIN--- 40  
1369 SEILFENSBEADEAEQNRVRKNSKPRRPDRKGNRNNRNDSTRRTDANHHNSVNYHH 1428  
41 -----GALV--PP--AAL-----I-----S-----I 52  
1429 GHNKPORQGA--TOOPKRSALHPENDNDTPSTPTVILTADPTTNPPCRQAPGQDNI 1487  
53 -----I-----Q-----KG-----L--Q-----Y---V 60  
1488 PHENDIPGSSSTTTQSSPPDDTNYSGPEHRNRIQNHRIIAQTNEFFGLYAYGNDIAV 1547  
61 -----E-----A-----EV-----S-----I---N 67  
1548 KKNSHKLMPPNEORLRSCEFQFMNGDSIYATENLEVIKTYNTSGEGYCGYNALKILYPN 1607  
68 -----E-----D-----99  
1608 LDLTLEEMQEIVGSETQFSDWEIMRVAQAKQLNLIVVTERCALVNKSVCSNEFGVICHCR 1667  
70 --G-----TL-----F-----D-----G-----R-----P--- 77  
1668 HRGVMLEHWEAALAIQKGFADYHPTFTNALTREDLLDFAKSSGLNKNKVSFVLGDPRL 1727  
78 -I-----ESLS-LI-D--A--VM-----P-----D-----D--- 90  
1728 KLQTEMHESLSKVVEDVNPAGFVKITKGSTHYVTNDRNLNGYSPQYCTFTAQITDNQEL 1787  
91 -----VVOT-----R-----Q-----QA-----98  
1788 VELLLLAYAPARVHTDWFYDRPESVPADATOGIHDYKQLVRDIAEVNQACAGDLIKS 1847  
99 -Y--R-D-K-----L-AQ-----Q-----H-----107  
1848 ELIQNRVDCKEYINNHLFKVQTKGLKPGDLISGLVGIQSTSYVDVSHVIGVIRTNQL 1907  
108 -----A-A-----A-----AAA---AAA---116  
1908 WCTCVTEKCAKATKIKVDLYRNKNTGSLRALFGIFRWFDRDITLLEKATAVDAIAGWG 1967  
117 -AT-----NO-----OG-----S-----123  
1968 KSTEIVKLVNDOCTVVAQSAASVNILEKLEQGGKQMKVMSIEKMTQQVNTPTLVLD 2027  
124 -A-----KN-----G-----E-----128  
2028 EASMITWETLSLTGQVENVLYGNTLOICVLDMYRTGSRATKSILOAGIIRHYTT 2087  
129 --N-----T-A--N-----G--BE-----135  
2088 HRIGNPLARELSLVTKELTNAKHETNFTCKSWDAVRWAELTISIAGSLBEPVILCFYNN 2147  
136 --N-----G--A-----H-----T-----140  
2148 VRVNMVLYKVCGRVDTIHKFQGLEADNVILWOCPTGPTGRITLDKHQCLSAATRAKN 2207  
141 --I-----ANN-----HT-----DM-----N-----149  
2208 LVWISVNEYSNNVPLHKRMGATIGGSKTQPTDENNLADKTLQVVSLLSKVMYISKDKT 2267  
150 -----E-----VD-----G-----153  
2268 IERLECSRLQSQBEGNLVMLKNTLTHLDVDVYSGNDHSLRQSGSKLNEKFKMSLSLFS 2327

154 ---D-V---EI-----157  
2328 KEPAPDHVSSSTLGLAEIVDOANANKDRINSQIMNIYVQKNLAISNGGSAIAAVTVPL 2387  
158 ---P---S---N-K---AVV---LR-----166  
2388 TAMLAQPGITKAECFDEPTETLTANILMGVNLKIQLDPRNDKVLILGSGLSRFVAD 2447  
167 ---GH---E-S---E-V-F-IC-----175  
2448 SSEIKQLYNOKPNCCHYVLYKYNFDETSVKKMMFVRLVIVNYFGVLESGLHSV 2507  
176 ---AW---N---P---V---SD-----182  
2508 TIENELKMQSGCSLGGWVIFHNKPIIIVGDPSTYNTGNSRTVRFNSKQSKCPILA 2567  
183 -L-L-V-SGS---GDS-TA--RI--W-----196  
2568 YLRLMNVDSTGVKLPYDPVPGRGDFDMFEGDNMAALVIERVLKGFWMMAAKLKNMLTN 2627  
197 ---GH---NL-----SE-----200  
2628 TNQDLNCFYDNNSPELFDLSKARLESNGFTNFNTLVNEFVGTRLEFNGFTNFNTLVN 2687  
201 -NS---T-SG---P-----T-----207  
2688 VNSMDLRCPILCTLDKRYLAALMDQDNLIVWTQISGKPIHTFAGVTRVSNKAFKTMVAVR 2747  
208 Q---LV-L---R---HC---I-R-EG-----218  
2748 QMLNRLEFKLTNTQLSRDKPMGAPNPRGPLLKFWAQHEADNDKVAVARKYKGTQQAAMM 2807  
219 ---G---Q-----D-----221  
2808 KLSRMVLTAWVNGSEYASINMQHNVTITSTNYPENNLGLVDLYVAQAIHNTGYT 2867  
222 ---V---P---S---N---K-D-V---228  
2868 SALYITNACTAVLGCWHWDFSCPPDDGWSRFLSTHDTITNLADIKSILDSIAKLD 2927  
229 ---T---S---L-D-W---NS-----235  
2928 KOELSEETAVLKEKESERIHQILRTEDSWYSESYQNVKGTIVVNSCGISTES 2987  
236 ---E-G---TL---LA---TGS-----Y---245  
2988 IEKIMETTGAMEFLMCPVTLTPSDNTTAHVGVTGSNNVKITYPGARNLTLPNPEIYCYM 3047  
246 ---DGF-----A---R-----250  
3048 KSGKRATVQSOTKCD-FDMISLHTTTVMASVQILKVTITRFEHAMPKALPIPPSPFS 3106  
251 ---I---W---T---KDG-----NL---A 259  
3107 HQSDIEKWVQYPRFTCNFGGSENSLIITKRLVSKALLAKNERALRHSNLKDLQA 3166  
260 ---S---T-L---G-----Q-H-----265  
3167 YGRSYLMTQVYTERVNVINKNGLDEBAVCIAALTMKHRINEQLEHTMGLMETFRVDT 3226  
266 ---K-G-P---I---F-----270  
3227 DSITDIVVCKNLEKVMISANTMQEKLGPASFTGLLDEIANTQRVFIDERTNQR 3286  
271 ---A-LKN-N-----KK-----277  
3287 VMSWDSNIKHSDESGIKKWAFTADNLDMWIGMIQYRKNVLCRKEDIPKTKACPGF 3346  
278 ---GN-----279  
3347 KQWDCAKAVAOKEARKTDPGARVLVLEPIPLIGNVDITRAAFTQGTMTKNENVRSE 3406

280 -F---I---LSA---GVD-K---T-TII-----292  
3407 ITFRARQSIIRAQTSRLRSYLDNWNMVKYSALEAG-DLKVMVRELSRIAIKSTSTP 3465  
293 -W---D---A-H---T---GEA---KQ---Q---F-----304  
3466 TWHLKEDLEAMLDVFNLTAEHLEITKLIGDAVSVRWQSKAEKKQNEYDFIKSGLGLST 3525  
305 ---P-----305  
3526 LNDHDLNIVLSLAVDRPSKVKKOLLTNTLGSQMSVNEAVMLRRKIROQWCEASNCQL 3585  
306 ---FH---S---A-----309  
3586 VDHHDHANSVSTTEQNOQKLNILSLHEAGICDINLSDASPLKIRASSYLPATTCVQ 3645  
310 ---P---ALDVD---W-----OS-----318  
3646 TPGMYELTSKPKVEIIEAMDNDCEVWRCIEKYVTQNIIBPHFRITNGLRAIMOQSKMLTES 3705  
319 ---N---N-T---F---AS-C---STD---328  
3706 QAIIVCOLLGLMCCLIQNGETGTVNFAPNKFVQLMRLSQDASVDHCVLINLLGADGVK 3765  
329 ---M---C---I---HV---C-----334  
3766 RLSPENIAKENMEQLEHVCVDNKPIVA GENPYACVSHVIDNEDLARLTSQSPYEKI 3825  
335 ---K---335  
3826 PVLDCRFEMLPVNNVMARMTKTSATLLKVTPTTGTGAGSLALTRNYYKATPMVTRPG 3885  
336 ---L---G-O-----338  
3886 DVLVLRDCEKKVMCTHTTIVDTFGVEQIILEITDDIGDGAUCLDLGLHCLTPADTRPR 3945  
339 -DRP---IKT-F---QG---H---T 349  
3946 RORPNTSNEVNSLNIETQRYLEATKGLRSHVGQEGKLIHYFDNRHHMYDEHDYIDA 4005  
350 ---N-E---VN---A---IK-W---D-P-----359  
4006 MLDENIVIELDPTVQVPTGVNIVNLICIRMFRLAMLYGIPTVWAYCDDNPSAAQLAWFL 4065  
360 ---360  
4066 KYTKMGKVFNNVTTEDLTNTFAMEIINALFSTYQHTTNTTNRVHLRKAETILSKEK 4125  
360 ---T-G---N---L---L-364  
4126 VSELEFNTTIDISVFTAVLMGLENEVLKSNQGTCTVFDLIPETPEPPQLQDDMLSGILV 4185  
365 ---A---S---CS---DD-M-----371  
4186 DIQCAVYHPTGTKVALRVVESLSHEIACSTKVGVDPKDDENRAGASNLTSVPASDK 4245  
372 ---T-----L---KI 375  
4246 ADPGSSQLSCEENMSQTGTQPVACVHIEPCDHAANASDDVLKEITQPEQNNHDIISKI 4305  
376 ---W-----S-----M-----378  
4306 LATAQAKSNPEDKSNKSGATSADLQAIQSLCSDPDGMSQVADQAKKLEVMNFSEVT 4365  
379 ---K-Q---D-----381  
4366 LSGIHRORNVNSOGKYLQVRDISERWAIYKIGHGSDVPTNPGWVWLAFFPMRKLK 4425  
382 ---NC-----V-H---D-----386  
4426 VYKEKEIKVTMVAGNCAVRLNRKIFREP KSVFTNVKPEVMDLILTYTSANNKOYPTN 4485  
387 LQ--A-----H-----N-----K-E-I-----Y-----395



Db 4486 LQRLADKYLHGRVSHVYVGGYNRDLQTCVELTKRMERVIVVTDAPDVNQLDADIQV 4545  
QY 396 ---TIK---W---SP---TG---P--- 404  
Db 4546 REKIITIKVTSQPEYPYRAIWMRFMILPANIDPECELYILANGREMAPNDMELVDRVMEESF 4605  
QY 405 ---G---T---N---N--- 408  
Db 4606 EDFDIVSALHCOAFNVCCGAKYNHAKFTDYNFNDIYEAGNSYGGDEIYMERYKVKMM 4665  
QY 409 ---P-NA---NML---SASP--- 420  
Db 4666 WLVIYILPGGAWYIGTEGKWLNVALDITICLBKEGERTRIDRVKGLFCFSAFFHVPDPTKG 4725  
QY 421 -D-STV--- 426  
Db 4726 IYDSTVDMSVDGIPWQYAMQINGHSVTENPVNRHSTGEGTGNPRNISNDGKFWQLNPN 4785  
QY 427 -W-D---V---D---R-G---I---C--- 434  
Db 4786 NLLDPLGKSEKWKIQFLDSSVKCIKESQSLPIQLGFLKPIVVPFANQACEATGVSM 4845  
QY 435 ---T---K---H-Q---E---P--- 438  
Db 4846 NNTGTLVVEVNIQAOKLAARNEPLQHOTLIKSHLTLMSTTGRHASKWPEYTHLMKADLQ 4905  
QY 439 ---T---K---H-Q---E---P--- 445  
Db 4906 NCQVYVEGSAITLQWKRETOHAQVYTBEDRVFVRIPOALDHSVGTATKLVKVTGNE 4965  
QY 446 -YS---V---A---F---S---P--- 452  
Db 4966 LGLSSSGVLARVLKNDRLAHHWLETOHLQKLEFDRRIISHESSGWTFKPRIGQT 5025  
QY 453 ---D---V---H---I---W--- 459  
Db 5026 LYLADVLTAAGIECVNHITTSREWKEALKNPLTKVGIMPLTNSVLVSGKLGCVQPIPR 5085  
QY 456 -Y---N---N---TG--- 474  
Db 5266 MKLTSRVPGQKINLKTLLTNRPQTRPVPTQVMGNGENAVTGRLGSLVPLRREPNNVT 5325  
QY 475 ---VH--- 476  
Db 5326 HELHKFTAYYRGWERNVLKDFKANTITISDADVKTWLSRRSDWKALATSTIKMLETGLP 5385  
QY 477 ---VH--- 478  
Db 5386 SNPMNAVNVHVKTESLLKANPIYWRQTQRIIVWQPKELCALMSPAFIAKRLKEVLR 5445  
QY 479 ---S---Y--- 480  
Db 5446 DEIVYTDGLTFPDLMSARARTIQDYVVFEDDLVIQDROTDQELIDLEFQVMDLGLDINLA 5505  
QY 481 ---R---G---TG--- 484  
Db 5506 NLWRLVHINKVRFKQHSWGOLDAMRLTGQATTALGNAINTLNVHSFVIEHROAIKLMFV 5565  
QY 485 -G---I---F---E---V---C--- 490

Db 5566 LGDDNITFMSAEPNLTKYKRLMSRYNMRSKPOVRNVGTFCILLSYRNSFGHCEVGPDF 5625  
QY 491 --- 492  
Db 5626 VLRLHRFEVTGGNHVELPVAVMPHNVRIGTKFFNGRYLHVHVNSEAEAEPEVLELRNH 5685  
QY 493 A---A---G---D---K---VCAS---A--- 502  
Db 5686 ANRSYILVLKTNSEWRLDGVQVHDLSDKIDTSGNLAEALGEGASGMGNPMPMAAGMSY 5745  
QY 503 ---S---D---G---S---VCV---L-D---L---RK 514  
Db 5746 LHMGLKSEADKINTEKEFGLERLKWYDVASCKNATCIKYNMSPMELEDNIGELLKMKIRK 5805  
  
RESULT 15  
S58870  
reelin precursor - mouse  
C;Species: Mus musculus (house mouse)  
C;Date: 06-Dec-1996 #sequence revision 06-Dec-1996 #text\_change 21-Jul-2003  
C;Accession: S58870; S71844; I49297  
R;D'Arcangelo, G.; Miao, G.G.; Chen, S.C.; Soares, H.D.; Morgan, J.I.; Curran, T.  
Nature 374, 719-723, 1995  
A;Title: A protein related to extracellular matrix proteins deleted in the mouse mutant  
A;Reference number: I49297; MUID:95231649; PMID:7715726  
A;Accession: S58870  
A;Status: nucleic acid sequence not shown  
A;Molecule type: mRNA  
A;Residues: 1-3461 <DAR>  
A;Cross-references: EMBL:U24703; NID:g902486; PID:g902487  
R;D'Arcangelo, G.  
submitted to the EMBL Data Library, April 1995  
A;Reference number: S71844  
A;Accession: S71844  
A;Molecule type: mRNA  
A;Residues: 1-215,'N',217-1905,'S',1907-3355,'V',3357-3391,'N',3393-3461 <DA2>  
A;Cross-references: EMBL:U24703; NID:g902486; PID:g902487  
F;1-27/Domain; signal sequence #status predicted <SIG>  
F;28-3461/Product; reelin #status predicted <MAT>  
F;1769-1795/Domain; EGF homology <EGF>

Query Match 70.3%; Score 2579.8; DB 2; Length 3461;  
Best Local Similarity 11.7%; Pred. No. 1.8e-52;  
Matches 390; Conservative 95; Mismatches 16; Indels 2842; Gaps 359;

QY 1 M-----SI-----S-----S-----D-----EV----- 8  
Db 1 MERGCWAPRALVLAVLLLATLRARAATGYPRFPFLCTHGHGEGDGEQGVLSL 60  
QY 9 ---N-----F---LV---VRY---LQ---ES-----GF---S-H 23  
Db 61 HIAGNPTVYVPGQYHVHTISTSTFFDGLLVTLG-YTSTSIQSSQSIGSSAFGFGIMSDH 119  
QY 24 ---S---A-----F-----T-----F-----G---I----- 30  
Db 120 QFGNQFMCVWASHVSHLPTNLSFWVIAPAGTGCWNFMATATHRGQVIFKDALAQQLC 179  
QY 31 ---E---SH-----I-----S---Q-----SNI-----N-----GAL----- 43  
Db 180 EQGAPTEATAYSHLAIHSDSVILRDDFDSDYQQLNPNINWECNCEMGECQGTIMHGN 239  
QY 44 -V---P---P-----A-AL---I-----SI-----IQ--- 54  
Db 240 AVTFCEPYGPRELTTTCLNTTASVLQFISIGSGSCRFSDPSITVSYAKNNTADMIQLE 299  
QY 55 ---KG---LQV---V-BA-E-----VGIN----- 67  
Db 300 KIRAPSNVSTVHILYLPBEAKGESVQFQWKQDSLVRGVEYACWALDNILV-INSARE 358  
QY 68 ---ED---GT-LF-----DG-----R----- 76  
Db 359 VVLEDNLDVPDGTGNLWFFPGATVTKHSCQSDGNSIYFPHNGESSEFNATTRDVLSTEDIQ 418

QY 77 -----P-----I-----ES-LSLI-----DA-----86  
Db 419 EQWSEFESOPTGWDILGAVVGADCGTVESGLSLVFLKDGKRLCTPYMDTTCYGNLRFY 478  
QY 87 -VM-----P-----D-----VV-----QT-----94  
Db 479 FVGGICDPCGVSHENDIILYAKIEGRKEHIALDTLTYSYKVPSPVSVINPELQTPATK 538  
QY 95 ---RQ---QAY-R---D-----K-LAQ-----Q-----106  
Db 539 FCLROKSHOQYNNRNVMAVDFHVLVPLVPTSMHMIQFSINLGCCTHQPGNSVSLEFSTNH 598  
QY 107 ---HA---A---AAA-AA-----AA-----AT---118  
Db 599 GRWSLLHTECLPEICAGPHLPSTVYSSYSGWNRITIPNAALTRDTRIRWQTGP 658  
QY 119 ---N-----Q-----Q-----GS123  
Db 659 ILGNMAIDNVYIGPSCFKFCGRGQCTRHGCKCDPGFGSPACEMASQTFPMFISESFGS 718  
QY 124 AK---N---G-E-----N-----129  
Db 719 ARLLSYHNFYSIRGAEVFCGVLASGKALVFNKQGRRLITSLDSSQSRLFTLRG 778  
QY 130 ---T---A---NG---E-E---N-----G-137  
Db 779 SKSVLSTCRAPDOPGEGVLLHYSDNGITWKLLEHYSYVNYHEPRIISVELPDDARQFGI 838  
QY 138 ---A-H---T---I---AN-----N-----H-145  
Db 839 QFRWQPYHSSQGEDVWAIDEIVMTSVLFPNSISLDFNLVEVTQSLGFLGNVQPYCGHD 898  
QY 146 ---T-D---M-----MEVDG-----D-V-E---156  
Db 899 WTLCTGDSKLASSMRYETQSMQI-GASYMIQFLVMGCGQKYTPHMDNQVLEYSANH 957  
QY 157 ---I---PSN---K---A---VVL-----R-----166  
Db 958 GLTWHLVQEBCLSPMSCQFTSASIVHASEFTQWRRTVVLPQKTWSGATRRFWSQSY 1017  
QY 167 ---GH-----E-SE-----V---P---173  
Db 1018 TAQDEWALDNIYIQOCPNCSHGSCDHGVCRCDOGYQTECHPEAALPTSIMSDPENP 1077  
QY 174 ---I---C-----A-----W-----177  
Db 1078 SSWEDMQEVIGGEVWKEQCGVSSGSSLYFSKAGKQLVSWDLDTSWVDVFQFYQI 1137  
QY 178 ---N-P---S-----SDL-----LV-----SG-----187  
Db 1138 GGESAAKNKPDREEGILLQYNSNGGIQWHLLEAMYFSDFSKPRFVYLELPAAGKTPCTR 1197  
QY 188 ---SG-----D---S-----T-----A-----R-IW-196  
Db 1198 FRWKPVFSGEDYQWAVDIIILSEKQKQVWPVNPPTLPQNFYKPAFDYPMNQMSVWL 1257  
QY 197 ---N-LSN-S---T-----S-G-----P-----T-QL---209  
Db 1258 MLANEGMAKNDSCATTPSAMVFGKSDGRFAVTRDITLKPYYVLPKLNIGCTQSFSST 1317  
QY 210 -V-LR-H---C-----IR-EG-----QOD-----V222  
Db 1318 APVLLQYSHDAGMSWFLKCECPFASAAKCEGNSRELSEPTVYTYG-DFEWTRITIAI 1376  
QY 223 P-----S-N-----K-----D-----227  
Db 1377 PRSLASKTRFRWIOESSQKNVPPFGLDGYVISEPCPYCSGHGDCISGVCFDLGYTA 1436  
QY 228 ---V-T---S-ID-----W-----NSEG-TL---L---A-241  
Db 1437 AQGTCVSNTPNHSEMFDRFEGKSLPLMYKITGGQVGTGCGTLN-DGRSLYFNLGKREAR 1495  
QY 242 ---T-----GS-----Y-D-G-----F248

Db 1496 TVPLDTRNLSLVQFYIQIGSKTSGITYITPRARYEGLVVQYSGNDNGILWHLRLDFMSF 1555  
QY 249 -----A-RW-----T---K-DG-256  
Db 1556 LEPQIISIDLPREAKTPATAFR-WWQPOHQKSHAQWALGDVLIGNVDSQTFQDKLDGS 1614  
QY 257 ---N-----LAS-T-L---G-Q---H-----KG-267  
Db 1615 IDLOANWYRIQGOVDIDCL-SMDTALIFTEIGNPRYAETWDFHVSSEFLOWENMGC 1673  
QY 268 ---PIF-A---L-----K-W---NKK-----G-----278  
Db 1674 SKP-FSAGHIGLOQYSLNNGKDWLVBECVPPTICGVHYTESSTYTSERFQNRRTVY 1732  
QY 279 ---N-----P-I---LSAGVDKT-TI---I---W-----DA-----295  
Db 1733 LPLATNSPRTRFWIQNTYTVGAD-SWADNVILASGCPMCMCGRIGCDGRCVCDRGFG 1791  
QY 296 ---H-----H-----T---GE-----299  
Db 1792 GPFQVVPVPLPSILKDDFNGNLHPDLWPEVYGAERGNLNGETIKSGTCLIFKGEGLRMLI 1851  
QY 300 ---AK-Q---QF-----P-----F-----H-----SA-----309  
Db 1852 SRDLCTNTMYQVQSLRFIAKGTPEPERSHSILLOFVSQGVVTHLMDDEFYPPQTRILFIN 1911  
QY 310 ---P---A---L-----D-----V-----D-W316  
Db 1912 VLPYGAQTNATFRFLWQPYNNGKKEEIIIDDFIIDGNLNNPVLILLDTDFGPREDNW 1971  
QY 317 -----Q-S-N-NT-----321  
Db 1972 FFYPGNGIGLYCPYSKGAPEDSAMVFSVNEVGEHSITRDLNVNENTIIQFEINVGCS 2031  
QY 322 -----F-A-----S-CST-----327  
Db 2032 TDSSADPVRLEFRDRFGATWHLPLCYHSSSLVSSLCSTHSPSTVYAGTQGWRE 2091  
QY 328 -----D-----D-----MC-IH-----VC334  
Db 2092 VVHFGKLHLCGSVFRWYQGFYAGPQVMTWIDNVYIQOCEMYGHCSCINGTKCIC 2151  
QY 335 -----K-L---GQ---DR-----P-----I-KT-----P-Q-G-347  
Db 2152 DPGYSGPTCKISTKNPDLKDDFEGQLESDRFLMMSGKPSKCGILSSGNLFFNEDGL 2211  
QY 348 -----H-----T-N-----EV-----352  
Db 2212 RMLVTRDLDLSHARFVQFFMRLCGKGVDPDRSQPVLLQVSLNGLSWLSLQBFLLFSNS 2271  
QY 353 NA---I---K-----W-----357  
Db 2272 NVGRYIALEMPKARSSTLRWQPSENGHFYSPWVIDQILIGNIGNTVLEDDPSTL 2331  
QY 358 D-----P---TG-----N---LL-----ASCS--D369  
Db 2332 DSRKWLHPGCTKMPVCGSTGDALVIEKASTRYVVTDDIANNEDSLQIDFAASCSTVD 2391  
QY 370 ---DM-----T---L-KI-----W-----376  
Db 2392 SCVAIELEYSDGLGSLWHLVLRDCLPTNVCECRYHLQRIILVSDTFNKRWTRITLPLPSYTR 2451  
QY 377 S-----M-----KO---DN---CV-----H-----D---386  
Db 2452 SQATRFHWQAPDFDKQWIDNVYIGDCLDMCGHRCVQSGCVCDQWGLYCDPE 2511  
QY 387 ---L-----Q-----A-H-----390  
Db 2512 ETSLPQLKDNFNAPSNQNLVTVSGKSLTVCGAVASGLAHFSGCSRLVTVDLNLT 2571  
QY 391 NKE-I---Y---T-----IKW-----S-PTG-----P404

```

Db 2572 NAEFIQFYPMYGLITPSNRNQGVLLEYSVNGGITWNLLMEIFYDQYKSP-GFVNILLPP 2630
QY 405 -----CT-----NNP-----N-A--NLMLA-SA-----SP-----D 421
Db 2631 DAKIATRRFWQ-FRHDGLDQNDWADINVLISGSADQRTVMLDTFSSAPVQHERSPAD 2689
QY 422 S-TV-R-----L-----W-----DVD-----R-----GI--C-IH-----TL--TKH- 441
Db 2690 AGFVGRIAPFEMFLEDKTSVNEWLFHD-DCTVERFCDSPDGVMLCGSHDGREVYAVT-HD 2747
QY 442 -----Q-----E-P-----V-YSAF-----S-----P-D----- 453
Db 2748 LPTENWIMQFKISVCKPEKIAQNIHVQFSTDFGVSWYLVQPCLPADPKCGSVSQ 2807
QY 454 -----G-R-Y-L-AS-G-----S-----F-----D- 463
Db 2808 PSVFPTEGWKRITYPLPSLTGNPVRFRFYQKYSVQWAIQNFYLGPCLDNCGHGDG 2867
QY 464 -K-CV-----H-----I-----W-----NT- 471
Db 2868 LKEQCICDPGSGPNCYLTHSLKTLKRFDFSEIKPDLWMSLEGGSTCTCGVLAENTA 2927
QY 472 -----Q-T-----GA-L-----V-----H-----S----- 479
Db 2928 LYFGSIVRQAITQDILRGAKFLQYWGRIQSENNTSCHRVPYCRKEGVLLDFSTDG 2987
QY 480 -----Y--R--G-----T-----G-----GI----- 486
Db 2988 WTLLEHMDFKYISVRHDYILLPEGALTNTTLRWQPFVISNGLVVGVCACAVGTGQH 3047
QY 487 -----F-----E-----V-----C-----W-----NA 493
Db 3048 SDWWSRNHPSQLVDTFDDSGSSHEENWSFYPNVARTAGFCGNPSFHYWPNKKKKTNA 3107
QY 494 -----A-----GD-----K----- 497
Db 3108 LSSRELIQGYMMQFKIVVGCEATSCGDLHVMLEYTKDARSQSWQLVQTCPLPSSNS 3167
QY 498 VG-----A-----SA-----S-----D----- 504
Db 3168 IGCSPQFHEATYNAVSSWKRTITQLPDHVSSATQFRMIQGEETEKOSWAIHDVY 3227
QY 505 -G-----S-----VC----- 508
Db 3228 IGEACPCLCSGHGYCTTGAVCICDESFGQDCSVFSHELPSYIKONFESARVTEANWETI 3287
QY 509 ---V-----LDL-R--K 514
Db 3288 QGGVIGGCGQLAPYAHGDSLVFNGCQIRQAATKPLDLTRASK 3330

```

Search completed: January 3, 2005, 15:29:45  
Job time : 114.667 secs

**This Page Blank (uspto)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 3, 2005, 14:55:11 ; Search time 102.333 Seconds  
(without alignments)  
2889.993 Million cell updates/sec

Title: US-09-987-701-12

Perfect score: 3669

Sequence: 1 MSISDEVNFLVRYLQESG.....GDKVGASADGVCVLDLRK 514

Scoring table: BLOSUM30

Gapop 1.0 , Gapext 0.1

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt\_02.\*

1: uniprot\_sprot.\*

2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3665	99.9	514	1 TBLR MOUSE	Q8bhj5 mus musculus
2	3622	98.7	514	1 TBLR HUMAN	Q9bzk7 homo sapien
3	3583.5	97.7	519	2 Q7S2M9	Q7szm9 xenopus lae
4	3577	97.5	522	2 Q6GFC6	Q6gpc6 xenopus lae
5	3397.9	92.6	527	1 TBLX MOUSE	Q9qxe7 mus musculus
6	3350.2	91.3	526	1 TBLX HUMAN	Q60907 homo sapien
7	3347.4	88.5	522	1 TBLX HUMAN	Q9bq87 homo sapien
8	2937.5	80.1	511	2 Q7Q371	Q7q371 anopheles g
9	2916.9	79.5	700	1 EBI DROME	Q95tj9 drosophila
10	2827.8	71.6	5636	2 Q96RW7	Q96rw7 homo sapien
11	2822.8	71.5	20925	2 Q8NXX1	Q8nxx1 trichoderma
12	2620.7	71.4	7073	2 AAP49011	Aap49011 sars coro
13	2618.9	71.4	7191	2 Q6AXA9	Q6axa09 alternaria
14	2618.9	71.4	7191	2 AAP78735	Aap78735 alternari
15	2612.9	71.2	4599	1 LR1B MOUSE	Q9j118 mus musculus
16	2612.1	71.2	5175	2 Q8U0J3	Q8u0j3 caenorhabdi
17	2611.2	71.2	34350	2 Q8W442	Q8w442 homo sapien
18	2610.8	71.2	5198	2 Q7P518	Q7p518 caenorhabdi
19	2610.4	71.1	7105	2 Q7PXX9	Q7pxw9 anopheles g
20	2610.1	71.1	4796	2 Q9NL88	Q9nl88 drosophila
21	2609.7	71.1	7073	1 R1AB CVHSA	P59641 h replicase
22	2609.7	71.1	7073	2 Q6JH48	Q6jh48 sars corona
23	2609.7	71.1	7073	2 Q6R7Y8	Q6r7y8 sars corona
24	2609.7	71.1	7073	2 Q6RCW7	Q6rcw7 sars corona
25	2609.7	71.1	7073	2 Q6RCX8	Q6rcx8 sars corona
26	2609.7	71.1	7073	2 Q6RCY9	Q6rcy9 sars corona
27	2609.7	71.1	7073	2 Q6RD00	Q6rd00 sars corona
28	2609.7	71.1	7073	2 Q6RD11	Q6rd11 sars corona
29	2609.7	71.1	7073	2 Q6RD22	Q6rd22 sars corona
30	2609.7	71.1	7073	2 Q6RD33	Q6rd33 sars corona
31	2609.7	71.1	7073	2 Q6RD44	Q6rd44 sars corona

Q6rd55 sars corona  
Q6rd66 sars corona  
Q6tpe9 sars corona  
Q6uzf1 sars corona  
Q6uzf5 sars corona  
Q6v586 sars corona  
Q6va80 sars corona  
Q6va91 sars corona  
Q6vaa2 sars corona  
Aar14802 sars coro  
Aar14806 sars coro  
Aar14810 sars coro  
Aap82978 sars coro  
Aar91584 sars coro

## ALIGNMENTS

RESULT 1  
TBLR MOUSE STANDARD; PRT; 514 AA.  
ID Q8BHJ5; Q8CEG4; Q8VEG3; Q9EQD4;  
AC Q8BHJ5; Q8CEG4; Q8VEG3; Q9EQD4;  
DT 05-JUL-2004 (Rel. 44, Created)  
DT 05-JUL-2004 (Rel. 44, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE F-box-like/WD-repeat protein TBLR1 (Nuclear corepressor/HDAC3  
DE complex subunit TBLR1) (TBLR1-related protein 1).  
GN Name=Tblr1; Synonyms=Iral;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/c;  
RA Zhang X., Dormady S., Basch R.;  
RT "Identification of four human cDNAs that are differentially expressed  
RT by early hematopoietic progenitors."  
RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.  
[2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Testis;  
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;  
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,  
RA Nikaido I., Oeato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,  
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,  
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,  
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Baissel K.W.,  
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,  
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,  
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,  
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,  
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,  
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,  
RA Maglott D.R., Mulcais L., Marchionni L., Mckenzie L., Miki H.,  
RA Nagahina T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,  
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,  
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,  
RA Sandelin A., Schneider C., Sempke C.A., Setou M., Shimada K.,  
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,  
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,  
RA Wilming L.G., Wymahaw-Boris A., Yanagisawa M., Yang I., Yang L.,  
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,  
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,  
RA Shiraki T., Waki K., Kawai K., Aizawa K., Arakawa T., Fukuda S.,  
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,  
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,  
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,  
RT Birney E., Hayashizaki Y.;  
RT "Analysis of the mouse transcriptome based on functional annotation of  
RT 60,770 full-length cDNAs."  
RL Nature 420:563-573 (2002).

[3]  
 SEQUENCE OF 314-514 FROM N.A.  
 TISSUE=Breast tumor;  
 MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 Hopkins R.F., Jordan R., Moore T., Max S.I., Wang J., Hsieh F.,  
 Diatchenko L., Marusina K., Farmer A.S., Rubin G.M., Hong L.,  
 Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 Raha S.A., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 Blakeley R.W., Touchan J.W., Green E.D., Dickinson M.C.,  
 Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,  
 Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 and mouse cDNA sequences";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 CC -1- FUNCTION: F-box-like protein involved in the recruitment of the  
 ubiquitin/19S proteasome complex to nuclear receptor-regulated  
 transcription units. Plays an essential role in transcription  
 activation mediated by nuclear receptors. Probably acts as  
 integral component of the N-Cor corepressor complex that mediates  
 the recruitment of the 19S proteasome complex, leading to the  
 subsequent proteasomal degradation of N-Cor complex, thereby  
 allowing cofactor exchange, and transcription activation (By  
 similarity).  
 CC -1- SUBUNIT: Component of the N-Cor repressor complex, at least  
 composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.  
 CC Probable component of some E3 ubiquitin ligase complex. Interacts  
 with histones H2B and H4 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- DOMAIN: The F-box-like domain is related to the F-box domain, and  
 apparently displays the same function as component of ubiquitin E3  
 ligase complexes (By similarity).  
 CC -1- SIMILARITY: Belongs to the WD-repeat EBI family.  
 CC -1- SIMILARITY: Contains 1 F-box-like domain.  
 CC -1- SIMILARITY: Contains 1 Lish domain.  
 CC -1- SIMILARITY: Contains 8 WD repeats.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC ENBL; AF268195; BAG44738.1; -  
 CC ENBL; AK029595; BAC26526.1; -  
 CC ENBL; AK033347; BAC28241.1; -  
 CC ENBL; AK036064; BAC29294.1; -  
 CC ENBL; BC018512; AAH18512.1; -  
 CC HSP; P16649; 1ERJ.  
 CC InterPro: IPR006594; Lish.  
 CC InterPro: IPR011047; Quin\_alc\_DH\_like.  
 CC InterPro: IPR001680; WD40..  
 CC Pfam: PF00400; WD40; 8.  
 CC PRINTS: PR00320; GPROTEINRPT.  
 CC ProDom: PD000018; WD40; 3.  
 CC SMART: SM00667; Lish; 1.  
 CC SMART: SM00320; WD40; 8.  
 CC PROSITE; PS00896; LISH; 1.  
 CC PROSITE; PS00678; WD\_REPEATS\_1; 4.  
 CC PROSITE; PS00882; WD\_REPEATS\_2; 6.  
 CC PROSITE; PS0294; WD\_REPEATS\_REGION; 1.  
 CC Activator; Nuclear protein; Repeat; Transcription regulation;  
 KW

KW Ub1 conjugation pathway; WD repeat.  
 FT DOMAIN 4 36 Lish.  
 FT REPEAT 41 86 F-box-like.  
 FT REPEAT 167 206 WD 1.  
 FT REPEAT 223 262 WD 2.  
 FT REPEAT 264 303 WD 3.  
 FT REPEAT 306 344 WD 4.  
 FT REPEAT 347 386 WD 5.  
 FT REPEAT 389 437 WD 6.  
 FT REPEAT 440 479 WD 7.  
 FT REPEAT 481 513 WD 8.  
 FT DOMAIN 108 117 Poly-Ala.  
 FT CONFLICT 185 185 A -> V (in Ref. 1).  
 FT CONFLICT 374 374 K -> R (in Ref. 2; BAC29294).  
 FT CONFLICT 480 480 Y -> C (in Ref. 3).  
 SQ SEQUENCE 514 AA; 55661 MW; 13BEC1C2C7F8BF14 CRC64;  
 Query Match 99.9%; Score 3665; DB 1; Length 514;  
 Best Local Similarity 99.8%; Pred. No. 1.4e-65;  
 Matches 513; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 MISSEDVNFVLYRYLOESGFSHSAFTGIESHSISQSNINGALVPPAALISIIKGLQVY 60  
 DB 1 MISSEDVNFVLYRYLOESGFSHSAFTGIESHSISQSNINGALVPPAALISIIKGLQVY 60  
 QY 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQOAAAAAAAAAATNQ 120  
 DB 61 EAEVSIINEDGTLFDGRPIESLSLIDAVMPDVVQTRQOAYRDKLAQOAAAAAAAAAATNQ 120  
 QY 121 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVFIQANPV 180  
 DB 121 QGSAXNGENTANGEANGAHTIANNHTDMMEVDGVEIPSNKAVVLRGHESEVFIQANPV 180  
 QY 181 SLLIVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWSGTL 240  
 DB 181 SLLIVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGQDVPSNKDVTSLDWSGTL 240  
 QY 241 ATGSDYGFARIWTKGNLSTLQHGKPIFALKNNKGNFILLSAGVDKTTIWDATGEA 300  
 DB 241 ATGSDYGFARIWTKGNLSTLQHGKPIFALKNNKGNFILLSAGVDKTTIWDATGEA 300  
 QY 301 KQGFPHSAPALDVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360  
 DB 301 KQGFPHSAPALDVDMQSNNTFASCTDMCIHVCKLGQDRPIKTFQGHNEVNAIKWDPT 360  
 QY 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTGNPNANMLASAF 420  
 DB 361 GNLLASCSDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPTGNPNANMLASAF 420  
 QY 421 DSTVRLWDVDRGICIHITLTKHQEPVYSVAFSPDGRYLASGSPDKCVHIWNTQTALVHSY 480  
 DB 421 DSTVRLWDVDRGICIHITLTKHQEPVYSVAFSPDGRYLASGSPDKCVHIWNTQTALVHSY 480  
 QY 481 RGTGGIFFEVCWNAAGDKVGSASDGSVCVLDLRK 514  
 DB 481 RGTGGIFFEVCWNAAGDKVGSASDGSVCVLDLRK 514  
 RESULT 2  
 ID TBLR HUMAN STANDARD; PRT; 514 AA.  
 AC Q9BZK7; Q9H211; Q9H9A1;  
 DT 28-FEB-2003 (Rel. 41, Created)  
 DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DT 01-OCT-2004 (Rel. 45, Last annotation update)  
 DE F-box-like/WD-repeat protein TBLR1 (Nuclear receptor corepressor/HDAC3  
 DE complex subunit TBLR1) (TBL1-related protein 1).  
 GN Name=TBLR1; Synonyms=IRAL;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]

SEQUENCE FROM N.A., IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR COMPLEX WITH NCOR1, NCOR2, GPS2, TBL1X AND HDAC3.  
MEDLINE=21930350; PubMed=11931768;  
Zhang J., Kalkum M., Chait B.T., Roeder R.G.;  
"The N-COR-HDAC3 nuclear receptor corepressor complex inhibits the JNK pathway through the integral subunit GPS2.";  
Mol. Cell 9:611-623 (2002).  
[2]  
SEQUENCE FROM N.A.  
PubMed=11063877;  
Zhang X., Dormady S.P., Basch R.S.;  
"Identification of four human cDNAs that are differentially expressed by early hematopoietic progenitors.";  
Exp. Hematol. 28:1286-1296 (2000).  
[3]  
SEQUENCE FROM N.A.  
PubMed=14702039; DOI=10.1038/ngl1285;  
Ota T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R., Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Makita H., Sekine M., Obayashi M., Nishi T., Shibahata T., Tanaka T., Ishii S., Yamamoto J.-I., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K., Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A., Sudo H., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M., Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y., Abe K., Kamihara K., Katsuta N., Sato K., Tanikawa M., Yamazaki M., Ninomiya K., Ishibashi T., Yamashita H., Murakawa K., Fujimori K., Tanai H., Kimata M., Watanabe M., Hiraoaka S., Chiba Y., Ishida S., Ono Y., Takiguchi S., Watanabe S., Yosida M., Hotuta T., Kusano J., Kanehori K., Tanaka H., Hara H., Tanase T.-O., Imose N., Musashino K., Yuuki H., Oshima A., Sasaki N., Aotsuka S., Yoshikawa Y., Matsunawa H., Ichihara T., Shiohata S., Sano S., Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O., Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H., Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B., Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y., Fujimori Y., Komiyama M., Tashiro H., Tanigami A., Fujiwara T., Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Ohmori Y., Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S., Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T., Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M., Togaishi T., Oyama K., Hata H., Watanabe M., Komatsu T., Mizushima-Sugano J., Satoh T., Shirai Y., Takahashi Y., Nakagawa K., Okumura K., Nagase T., Nomura N., Kikuchi H., Masuho Y., Yamashita R., Nakai K., Yada T., Nakamura Y., Ohara O., Isogai T., Sugano S.;  
"Complete sequencing and characterization of 21,243 full-length human cDNAs.";  
Nat. Genet. 36:40-45 (2004).  
[4]  
SEQUENCE OF 1-125 FROM N.A.  
RC TISSUE=Adrenal cortex;  
MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Murray D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Heiton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalek U., Smalhus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";  
Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
[5]

RP COMPONENT OF THE N-COR COMPLEX WITH TBL1X, CORO2A AND HDAC3, AND  
RX HISTONE-BINDING.  
PubMed=12628926;  
Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,  
Wong J.;  
"Purification and functional characterization of the human N-COR complex: the roles of HDAC3, TBL1 and TBL1L1.";  
EMBO J. 22:1336-1346 (2003).  
[6]  
FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.  
RP PubMed=14980219;  
Rixi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;  
"A corepressor/coactivator exchange complex required for transcriptional activation by nuclear receptors and other regulated transcription factors.";  
Cell 116:511-526 (2004).  
CC -!- FUNCTION: F-box-like protein involved in the recruitment of the ubiquitin/19S proteasome complex to nuclear receptor-regulated transcription units. Plays an essential role in transcription activation mediated by nuclear receptors. Probably acts as integral component of the N-COR corepressor complex that mediates the recruitment of the 19S proteasome complex, leading to the subsequent proteasomal degradation of N-COR complex, thereby allowing cofactor exchange, and transcription activation.  
CC -!- SUBUNIT: Component of the N-COR repressor complex, at least composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.  
CC Probable component of some E3 ubiquitin ligase complex. Interacts with histones H2B and H4.  
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
CC -!- TISSUE SPECIFICITY: Ubiquitous.  
CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and apparently displays the same function as component of ubiquitin E3 ligase complexes (By similarity).  
CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.  
CC -!- SIMILARITY: Contains 1 F-box-like domain.  
CC -!- SIMILARITY: Contains 1 Lish domain.  
CC -!- SIMILARITY: Contains 8 WD repeats.  
CC -!- CAUTION: Ref.4 sequence differs from that shown due to a potential poly-A sequence starting in position 125.  
-----  
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
EMBL; AF314544; AAK00301.1; -;  
EMBL; AF268193; AAG44736.1; -;  
EMBL; AK022956; BAB1431.1; -;  
EMBL; BC060320; AAH60320.1; ALT\_TERM.  
DR HSSP; P16649; IERJ.  
DR MIM; 608628; -;  
DR InterPro; IPR001047; Lish.  
DR InterPro; IPR001680; WD40.  
DR Pfam; PF00400; WD40; 8.  
DR PRINTS; PR00320; GPROTEINBRPT.  
DR PRODOM; PD000018; WD40; 3.  
DR SMART; SM00667; Lish; 1.  
DR SMART; SM00320; WD40; 8.  
DR PROSITE; PS00896; LISH; 1.  
DR PROSITE; PS00678; WD\_REPEATS\_1; 4.  
DR PROSITE; PS00882; WD\_REPEATS\_2; 6.  
DR PROSITE; PS00294; WD\_REPEATS\_REGION; 1.  
KW Activator; Nuclear protein; Repeat; Transcription regulation;  
KW Ub1 conjugation pathway; WD repeat.  
FT DOMAIN 4 36 Lish.  
FT DOMAIN 41 86 F-box-like.  
FT REPEAT 167 206 WD 1.  
FT REPEAT 223 262 WD 2.  
FT REPEAT 264 303 WD 3.





```

RC TISSUE=Embryo;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Siemmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buettow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Usdin T.B., Tohiyuki S., Carninci P., Prange C.,
RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Vallalon D.K., Muzny D.C., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX MEDLINE=22341132; PubMed=12454917;
RA Klein S.L., Strausberg R.L., Wagner L., Pontius J., Clifton S.W.,
RA Richardson P.;
RT "Genetic and genomic tools for Xenopus research: The NIH Xenopus
Initiative.";
RL Dev. Dyn. 225:384-391(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RX Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.
CC -I- SIMILARITY: Contains 8 WD repeats.
DR EMBL; BC073215; AAH73215.1; -.
DR InterPro; IPR006594; Lish.
DR InterPro; IPR011047; Quin_alc_DH_like.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINBRPT.
DR ProDom; PD000018; WD40; 3.
DR SMART; SM00667; Lish; 1.
DR SMART; SM00320; WD40; 8.
DR PROSITE; PS00896; Lish; 1.
DR PROSITE; PS00678; WD REPEATS 1; 4.
DR PROSITE; PS00082; WD REPEATS 2; 6.
DR PROSITE; PS0294; WD REPEATS REGION; 1.
KW Hypothetical protein; Repeat; WD repeat.
SQ SEQUENCE 522 AA; 56284 MW; 4FB0F58F503B461F CRC64;

Query Match 97.5%; Score 3577; DB 2; Length 522;
Best Local Similarity 96.0%; Pred. No. 1.le-63;
Matches 502; Conservative 4; Mismatches 7; Indels 10; Gaps 4;

QY 1 MSISSEVNFVLYRYLQESGFSATFTGIESHSQSNINGALVPPAALISIIQGLQYV 60
DB |||||
DB 1 MSISSEVNFVLYRYLQESGFSATFTGIESHSQSNINGALVPPAALISIIQGLQYV 60
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAVPDVDTQQAQRDKLAQOH-----AAAAAAA 115
DB |||||
DB 61 EAEVSINEDGTLFDGRPIESLSLIDAVPDVDTQQAQRDKLAQOH-----AAAAAAA 120
QY 116 AAT---NQGSAGKNGENTANGEAGHTIANNHTDMWEVDGDVEI-PSNKAVVLRGHESE 171
DB |||||
DB 121 AATSAQQPPAKNGSTANGENGAGHALANNHTDMWEVDGDVEIIPPS-KAVVLRGHESE 179
QY 172 VFTCAWNPVSDLLVSGSGDSTARIWNLSENSTSGTQLVLRHCIRGGQDVPSPKNDVTSL 231
DB |||||
DB 180 VFTCAWNPVSDLLVSGSGDSTARIWNLSENSTSGTQLVLRHCIRGGQDVPSPKNDVTSL 239

232 DWNSEGTLLATGSGYDGFARIWTKDGNLSTLQGHKGPIFALKWKNKGNFILSAGVDKTTI 291
240 DWNSEGTLLATGSGYDGFARIWTKDGNLSTLQGHKGPIFALKWKNKGNFILSAGVDKTTI 299
292 IWDHTGEAKQOPPHSAPALDWDVQSNNTFASCSTDMCIHVCKLGDGDRPKTFQGHNTNE 351
300 IWDHTGEAKQOPPHSAPALDWDVQSNNTFASCSTDMCIHVCKLGDGDRPKTFQGHNTNE 359
352 VNAIKWDPPTGNLLASGSDMTLKIWSMKDNCVHDIQAHNKEIYTIKWSPGPGTNNPNA 411
360 VNAIKWDPPTGNLLASGSDMTLKIWSMKDNCVHDIQAHNKEIYTIKWSPGPGTNNPNA 419
412 NLMLASASFDSTVRLWVDVDRGICHTLTHTKHQEPVYSVAFSPDGRYLASGSDFCVHIWNT 471
420 NLMLASASFDSTVRLWVDVDRGICHTLTHTKHQEPVYSVAFSPDGRYLASGSDFCVHIWNT 479
472 QTGALVHSYRGTCGTFEVCWNAAGDKVGSASDGSVCVLDLRK 514
480 QTGALVHSYRGTCGTFEVCWNAAGDKVGSASDGSVCVLDLRK 522

RESULT 5
TBXL_MOUSE STANDARD; PRT; 527 AA.
AC Q9QXE7; Q8BMM0; Q8BYQ4; Q8COA1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein TBXLX (transducin beta-like IX protein).
GN Name=TBlix; Synonyms=TBll;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RC TISSUE=Hypothalamus, Medulla oblongata, and pituitary;
RX MEDLINE=22354683; PubMed=12466851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Hono H., Kondo S.,
RA Nikaido I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C., Gojobori T.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanapin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bradt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.F., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M., King B.L.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reid D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Semple C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita M.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wynnshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavalon N., Zhu Y., Zimmer A., Carninci P., Hayashizaki Y.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Hayashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

```

Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marudina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Udén T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Boesak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahney J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.; et al.  
 "Generation and initial analysis of more than 15,000 full-length human RT and mouse cDNA sequences";  
 Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [3]  
 SEQUENCE OF 200-527 FROM N.A.  
 STRAIN-129/SwJ;  
 Botcherby M.R.M., Straw R., Clarke D., Greyststrong J.S., Weston P., Hunter G., Kimberly C., Rhodes M.;  
 RA Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 TISSUE SPECIFICITY.  
 RP MEDLINE=9264241; PubMed=10330347;  
 RX Bassi M.T., Ramas R.S., Caciotti B., Winship I.M., De Grandi A., Riboldi M., Townes P.L., Beighton P., Ballabio A., Borsani G.;  
 RA "X-linked late-onset sensorineural deafness caused by a deletion  
 RT involving Oai and a novel gene containing WD-40 repeats.";  
 RL Am. J. Hum. Genet. 64:1604-1616(1999).  
 CC -1- FUNCTION: F-box-like protein involved in the recruitment of the  
 CC ubiquitin/19S proteasome complex to nuclear receptor-regulated  
 CC transcription units. Plays an essential role in transcription  
 CC activation mediated by nuclear receptors. Probably acts as  
 CC integral component of corepressor complexes that mediates the  
 CC recruitment of the 19S proteasome complex, leading to the  
 CC subsequent proteasomal degradation of transcription repressor  
 CC complexes, thereby allowing cofactor exchange (By similarity).  
 CC -1- SUBUNIT: Component of the N-CoR repressor complex, at least  
 CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.  
 CC Component of a E3 ubiquitin ligase complex containing UBE2D1,  
 CC STAH1, CACVBP/SIP, Skp1A, APC and TBL1X. Probably part of other  
 CC corepressor complexes, that do not contain NCOR1 and NCOR2.  
 CC Interacts with histones H2B, H3a and H4 (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed in the cochlea.  
 CC -1- DOMAIN: The F-box-like domain is related to the F-box domain, and  
 CC apparently displays the same function as component of ubiquitin E3  
 CC ligase complexes (By similarity).  
 CC -1- SIMILARITY: Belongs to the WD-repeat EBI family.  
 CC -1- SIMILARITY: Contains 1 F-box-like domain.  
 CC -1- SIMILARITY: Contains 1 Lish domain.  
 CC -1- SIMILARITY: Contains 8 WD repeats.  
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to erroneous  
 CC gene model prediction.  
 -----  
 This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----  
 DR EMBL; AK030547; BAC27015.1; -;  
 DR EMBL; AK031937; BAC27612.1; -;  
 DR EMBL; AK038674; BAC30092.1; -;  
 DR EMBL; BC043105; AB043105.1; -;  
 DR EMBL; F38006; CAB61534.1; ALT\_SEQ.  
 DR MGD; MGII:1336172; Tbl1x.

DR InterPro; IPR001680; WD40.  
 DR Pfam; PF00400; WD40; 5.  
 DR PRINTS; PR00320; GPROTEINRPT.  
 DR PRODOM; PD000018; WD40; 2.  
 DR SMART; SM00320; WD40; 5.  
 DR PROSITE; PS00896; LISH; 1.  
 DR PROSITE; PS00678; WD\_REPEATS\_1; 4.  
 DR PROSITE; PS00082; WD\_REPEATS\_2; 6.  
 DR PROSITE; PS0294; WD\_REPEATS\_REGION; 1.  
 KW Activator; Nuclear protein; Repeat; Transcription regulation;  
 KW Ub1 conjugation pathway; WD repeat.  
 FT DOMAIN 41 36  
 FT DOMAIN 41 86 F-box-like.  
 FT REPEAT 180 219 WD 1.  
 FT REPEAT 236 275 WD 2.  
 FT REPEAT 277 316 WD 3.  
 FT REPEAT 319 359 WD 4.  
 FT REPEAT 360 399 WD 5.  
 FT REPEAT 402 450 WD 6.  
 FT REPEAT 453 492 WD 7.  
 FT REPEAT 494 526 WD 8.  
 FT DOMAIN 108 121 Poly-Ala.  
 FT DOMAIN 130 133 Poly-Ala.  
 FT CONFLICT 104 104 A -> T (in Ref. 1; BAC27015).  
 FT CONFLICT 349 349 L -> H (in Ref. 1; BAC27612).  
 SQ SEQUENCE 527 AA; 56802 MW; 146435A9C51DFDA6 CRC64;  
 Query Match 92.8%; Score 3397.9; DB 1; Length 527;  
 Best Local Similarity 87.6%; Pred. No. 6.9e-60;  
 Matches 465; Conservative 28; Mismatches 17; Indels 21; Gaps 9;  
 QY 1 MSTSSDEVNFLVYRYLQESGFSHSAPTGGIESHISOSNINGALVPPAALISIIKGLQYV 60  
 DB 1 MSTSSDEVNFLVYRYLQESGFSHSAPTGGIESHISOSNINGALVPPAALISIIKGLQYV 60  
 QY 61 EAEVSNEDGTLFDGRIPIESLSLIDAVMPDVVQTRQOAYRDKLAQOH--AAAAAAAAT 118  
 DB 61 EAEVSNEDGTLFDGRIPIESLSLIDAVMPDVVQTRQOAYRDKLAQOH--AAAAAAAAT 120  
 QY 119 -----NQGSAKNGENTANGEANGAHTIANNHDM-MEVDGDVEI-PSNKAVV 164  
 DB 121 ATSTAATTPAAAQQNPFPKNGEATVNGEENGAI--NNHS-KPMEIDGDEIPEPS-KATV 177  
 QY 165 LRGHSEVEFICAWNPNVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGDVPS 224  
 DB 178 LRGHSEVEFICAWNPNVSDLLVSGSGDSTARINWLSNSTSGPTQLVLRHCIREGGDVPS 237  
 QY 225 NKDVTSLDWNSBGTLLATSGYDGFARITWKGKGLASTLQHGKGPFPALKWNKKNFILSA 284  
 DB 238 NKDVTSLDWNSBGTLLATSGYDGFARITWKGKGLASTLQHGKGPFPALKWNKKNFILSA 297  
 QY 285 GVDKTTIIWDAHTGEAKQPPHSPALPDVDWQSN--TFASCTSDMCIHVCKLGQDRPIK 343  
 DB 298 GVDKTTIIWDAHTGEAKQPPHSPALPDVDWQ--NNTTFASCTSDMCIHVCKLGQDRPVK 356  
 QY 344 TFQGHTEVNAIKWDPDPTGNLLASCSDMTLKTWSMKODNCVHDLOAHNKEIYIKWSPG 403  
 DB 357 TFQGHTEVNAIKWDPDPTGNLLASCSDMTLKTWSMKODNCVHDLOAHNKEIYIKWSPG 416  
 QY 404 PGTNNPNANMLASAFDSTVRLWDVDRGICHTHTLTKHQEPVYSVAFSPDGRYLAGSGFD 463  
 DB 417 PATSPNSNIMLASAFDSTVRLWDVDRGICHTHTLTKHQEPVYSVAFSPDGRYLAGSGFD 476  
 QY 464 KCVHTWNTGTGALVHSYRGTTGFEVCWNAAGDKVGASADSGSVCLDLRK 514  
 DB 477 KCVHTWNTGTGALVHSYRGTTGFEVCWNAAGDKVGASADSGSVCLDLRK 527  
 RESULT 6  
 TELX HUMAN  
 ID TELX HUMAN STANDARD; PRT; 526 AA.  
 AC O60907; O86UY2;  
 DT 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)  
 DE 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE F-box-like/WD-repeat protein TBL1X (Transducin beta-like 1X protein)  
 DE (Transducin-beta-like 1, X-linked) (SNAP55).  
 GN Name=TBL1X; Synonyms=TBL1;  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A., TISSUE SPECIFICITY, AND DISEASE.  
 RX MEDLINE=99264241; PubMed=10330347;  
 RA Bassi M.T., Ramesar R.S., Caciotti B., Winship I.M., De Grandi A.,  
 RA Riboni M., Townes P.L., Beighton P., Ballabio A., Borsani G.;  
 RT "X-linked late-onset sensorineural deafness caused by a deletion  
 RT involving OAI and a novel gene containing WD-40 repeats";  
 RL Am. J. Hum. Genet. 64:1604-1616(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Lymph, and Pancreas;  
 RA MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Udwin T.B., Toehiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek A.M., Gunaratne P.H.,  
 RA Richards D.K., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzyzinski M.I., Skaleka U., Smalhus D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP IDENTIFICATION BY MASS SPECTROMETRY, AND COMPONENT OF THE N-COR  
 RP COMPLEX WITH NCOR2 AND HDAC3.  
 RX PubMed=10809664;  
 RA Guenther M.G., Lane W.S., Fischle W., Verdin E., Lazar M.A.,  
 RA Shiekhattar R.;  
 RT "A core SMRT corepressor complex containing HDAC3 and TBL1, a WD40-  
 RT repeat protein linked to deafness";  
 RL Genes Dev. 14:1048-1057(2000).  
 RN [4]  
 RP COMPONENT OF THE N-COR COMPLEX WITH NCOR2 AND HDAC3.  
 RX PubMed=10944117;  
 RA Li J., Wang J., Wang J., Nawaz Z., Liu J.M., Qin J., Wong J.;  
 RT "Both corepressor proteins SMRT and N-CoR exist in large protein  
 RT complexes containing HDAC3";  
 RL EMBO J. 19:4342-4350(2000).  
 RN [5]  
 RP SUBUNIT OF A COMPLEX WITH UBE2D1, CACYBP, SIAH1 AND APC.  
 RX PubMed=11389839;  
 RA Matsuzawa S.-I., Reed J.C.;  
 RA "Siah-1, SIP, and Ebi collaborate in a novel pathway for beta-catenin  
 RT degradation linked to p53 responses";  
 RL Mol. Cell 7:915-926(2001).  
 RN [6]  
 RP COMPONENT OF THE N-COR COMPLEX WITH NCOR1, NCOR2, GPS2, TBL1R AND  
 RP HDAC3.  
 RX MEDLINE=21930350; PubMed=119311768;  
 RA Zhang J., Kalkum M., Chait B.T., Roeder R.G.;  
 RT "The N-Cor-HDAC3 nuclear receptor corepressor complex inhibits the JNK  
 RT pathway through the integral subunit GPS2";  
 RL Mol. Cell 9:611-623(2002).  
 RN [7]

RP COMPONENT OF THE N-COR COMPLEX WITH TBL1R, CORO2A AND HDAC3, AND  
 RP HISTONE-BINDING.  
 RX PubMed=12628926;  
 RA Yoon H.-G., Chan D.W., Huang Z.-Q., Li J., Fondell J.D., Qin J.,  
 RA Wong J.;  
 RT "Purification and functional characterization of the human N-CoR  
 RT complex: the roles of HDAC3, TBL1 and TBL1R";  
 RL EMBO J. 22:1336-1346(2003).  
 RN [8]  
 RP FUNCTION, AND RECRUITMENT OF 19S PROTEASOME COMPLEX.  
 RX PubMed=14980219;  
 RA Perissi V., Aggarwal A., Glass C.K., Rose D.W., Rosenfeld M.G.;  
 RT "A corepressor/coactivator exchange complex required for  
 RT transcriptional activation by nuclear receptors and other regulated  
 RT transcription factors";  
 RL Cell 116:511-526(2004).  
 CC -!- FUNCTION: F-box-like protein involved in the recruitment of the  
 CC ubiquitin/19S proteasome complex to nuclear receptor-regulated  
 CC transcription units. Plays an essential role in transcription  
 CC activation mediated by nuclear receptors. Probably acts as  
 CC integral component of corepressor complexes that mediates the  
 CC recruitment of the 19S proteasome complex, leading to the  
 CC subsequent proteosomal degradation of transcription repressor  
 CC complexes, thereby allowing cofactor exchange.  
 CC -!- SUBUNIT: Component of the N-CoR repressor complex, at least  
 CC composed of NCOR1, NCOR2, HDAC3, TBL1X, TBL1R, CORO2A and GPS2.  
 CC Component of a E3 ubiquitin ligase complex containing UBE2D1,  
 CC SIAH1, CACYBP/SIP, SKP1A, APC and TBL1X. Probably part of other  
 CC corepressor complexes, that do not contain NCOR1 and NCOR2.  
 CC Interacts with histones H2B, H3a and H4.  
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).  
 CC -!- TISSUE SPECIFICITY: Ubiquitous.  
 CC -!- DOMAIN: The F-box-like domain is related to the F-box domain, and  
 CC apparently displays the same function as component of ubiquitin E3  
 CC ligase complexes (By similarity).  
 CC -!- DISEASE: Defects in TBL1X may be a cause of late-onset  
 CC sensorineural deafness (OASD) [MIM:300650]. OASD is an X-linked  
 CC recessive disorder characterized by ocular albinism and  
 CC progressive sensorineural hearing loss in the fourth and fifth  
 CC decades of life. OASD may be caused by deletion of both GPR143/OAI  
 CC and TBL1X adjacent genes; TBL1X defects possibly causing the  
 CC hearing phenotype.  
 CC -!- SIMILARITY: Belongs to the WD-repeat EBI family.  
 CC -!- SIMILARITY: Contains 1 F-box-like domain.  
 CC -!- SIMILARITY: Contains 1 Lish domain.  
 CC -!- SIMILARITY: Contains 8 WD repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Y12781; CAA73319.1; ALT INIT.  
 DR EMBL; BC032708; AAH32708.1; --\_INIT.  
 DR EMBL; BC052304; AAH52304.1; ALT\_INIT.  
 DR HSSP; FL6649; IERJ.  
 DR GENE; HGNC:11585; TBL1X.  
 DR MIM; 300196; --  
 DR MIM; 300650; --  
 DR GO; GO:0007605; P:perception of sound; TAS.  
 DR GO; GO:0007165; P:signal transduction; TAS.  
 DR GO; GO:0007601; P:visual perception; TAS.  
 DR InterPro; IPR006594; Lish.  
 DR InterPro; IPR011047; Quin\_alc\_DH\_like.  
 DR InterPro; IPR001680; WD40-  
 DR Pfam; PF00400; WD40; 8.  
 DR PRINTS; PR00320; GPROTEINBRPT.  
 DR ProDom; PD000018; WD40; 3.  
 DR SMART; SM00667; Lish; 1.  
 DR SMART; SM00320; WD40; 8.

```
DR PROSITE; PS50896; LISH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 4.
DR PROSITE; PS50082; WD_REPEATS_2; 6.
DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
KW Activator; Deafness; Nuclear protein; Repeat;
KW Transcription regulation; Ub1 conjugation pathway; WD repeat.
FT DOMAIN 4 36 Lish.
FT DOMAIN 41 86 F-box-like.
FT REPEAT 179 218 WD 1.
FT REPEAT 235 274 WD 2.
FT REPEAT 276 315 WD 3.
FT REPEAT 318 358 WD 4.
FT REPEAT 359 398 WD 5.
FT REPEAT 401 449 WD 6.
FT REPEAT 452 491 WD 7.
FT REPEAT 493 525 WD 8.
FT DOMAIN 108 124 Poly-Ala.
SQ SEQUENCE 526 AA; 57048 MW; 98922F88EC42F6E9 CRC64;

Query Match          91.3%; Score 3350.2; DB 1; Length 526;
Best Local Similarity 86.2%; Pred. No. 7.1e-59;
Matches 456; Conservative 36; Mismatches 19; Indels 18; Gaps 10;

QY 1 MSISDEVNPLVRYLQESGFSAFTFGIESHSQSNINGALVPPAALISIIKGLQV 60
DB 1 MSISDEVNPLVRYLQESGFSAFTFGIESHSQSNINGTLVPPAALISIIKGLQV 60
QY 61 EAEVINEDEGTLPDGRPIESLSLIDAMPDVVOTROQAVRDKLAQH--AAAAAAA--- 115
DB 61 EAEVINEDEGTLPDGRPIESLSLIDAMPDVVOTROQAFREKLAQQAASAAAAATAA 120
QY 116 --RATNQ-QG-S---AKGENTANGENGAAHTIANNHTDM-MEVDGDVEIPSKAVLR 166
DB 121 ATAATTSAGVSHQNPSPKREATVNGEENRAHSV--NNHA-KPEIDGVEIPSKAVLR 178
QY 167 GHESEVFIICAWNPSVOLLVSGSGDSTARINWLNSENSTSGPTQLVLRHCIREGGQDVPNSK 226
DB 179 GHESEVFIICAWNPSVOLLVSGSGDSTARINWLNSENSTSGPTQLVLRHCIREGGHVPNSK 238
QY 227 DVTSLDWNSEGTLLATGSDGPFARIWTKDGNLASTLQHKGPFPALKNKGNFIISAGV 286
DB 239 DVTSLDWNSEGTLLATGSDGPFARIWTKDGNLASTLQHKGPFPALKNKGNFIISAGV 298
QY 287 DKTITLWDATGKAQFPFHPSPALDVMQSN--TFASCSTDMCTHVCKLGDRPIKTF 345
DB 299 DKTITLWDATGKAQFPFHPSPALDVMQ--NTTFASCSTDMCTHVCKLGCDRPIKTF 357
QY 346 QGHTNEVNAIKWDPGNTGLLASCDDMTLKIWSKQDNCVHDLOAHNKEIYTIKWSPTGP 405
DB 358 QGHTNEVNAIKWDPGNTGLLASCDDMTLKIWSKQEVCIHDLQAHNKEIYTIKWSPTGPA 417
QY 406 TNNPNANMLASAFDSTVRLVDVDRGICHTLTKQEPYVAFSPDGRYLASGSFDC 465
DB 418 TSNPNSNIMLASAFDSTVRLVDIERGVCTHTLTKQEPYVAFSPDGRYLASGSFDC 477
QY 466 VHTWNTOTGALVSYRGTTGIFECVWNAAGDKVGASASGVCVLDLRK 514
DB 478 VHTWNTOSGNLVHSYRGTTGIFECVWNAAGDKVGASASGVCVLDLRK 526

RESULT 7
TBLY HUMAN
ID TBLY HUMAN STANDARD; PRT; 522 AA.
AC Q9BQ87;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein Tbl1y (Transducin beta-like 1Y protein)
DE (Transducin-beta-like 1, Y-linked).
GN Name=TBLY1; Synonyms=TB11;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```

SQ SEQUENCE 522 AA; 5668 MW; 4E020216422442D8 CRC64;
Query Match 88.5%; Score 3247.4; DB 1; Length 522;
Best Local Similarity 84.4%; Pred. No. 1.1e-56;
Matches 443; Conservative 36; Mismatches 30; Indels 16; Gaps 10;

QY 1 MSISDEVNPLVRYLOESGFSHSAFTGIESHSOSNINGALVPPAALISIIKGLQYV 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 MSISDEVNPLVRYLOESGFSHSAFTGIESHSOSNINGALVPPAALISIIKGLQYV 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAMPDVVOTROQAVYDKLAQQA-AAA--A-AAA- 115
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 EAEISINKDGTVDSPRIESLSLIVAVIPDVQMRQAFGEKUTQQQASAAATEASAMAK 120
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 116 AAT-----NQQSAKNGENTANGEANGHTIANNHTDM-MEYVDGVEIPSNKAVLGRH 168
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 AATWTPAAISQQPNKREATVNGEANGAHEI-NNHS-KPEWIDGDVEIPPNKATVLRGH 178
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 169 ESEVFTICAMNPVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIRREGQDVPSNKDV 228
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 179 ESEVFTICAMNPVSDLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIRREGGHDVPSNKDV 238
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 229 TSLDWNSECTLLATGSDYGFARITWTKDGNLASTLGQHKGPFIKWKKNKGNFILSAGVDK 288
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 239 TSLDWNSECTLLATGSDYGFARITWTKDGNLASTLGQHKGPFIKWKKNKGNFILSAGVDK 298
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 289 TTIIDAHTEGAQKQPFHSPALPDVDWQSN--TFASCSTDMCIHVCKLGODRPIKTFQG 347
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 299 TTIIDAHTEGAQKQPFHSPALPDVDWQ--NNMTFASCSTDMCIHVCKLGODRPIKTFQG 357
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 348 HTNEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNCVHDLQAHNKEIYTIKWSPTGPGTN 407
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 358 HTNEVNAIKWDPSGMLLASCDDMTLKIWSMKQDACVHDLQAHNKEIYTIKWSPTGPGTN 417
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 408 NPANILMLASASPDSTVRLWDVDRGICHTLTKHQPVPVSVAPSPGRVILASGSPKCVH 467
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 418 NPNSIMLASASPDSTVRLWDVDRGICHTLTKHQPVPVSVAPSPGRVILASGSPKCVH 477
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 468 IWNTOGALVHSYRGTTGGIPEVCWNAAGDKVGASASDGSVCVLDL 512
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 478 IWNTOGALVHSYRGTTGGIPEVCWNAAGDKVGASASDGSVCVLDL 522
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 8
Q70371 PRELIMINARY; PRT; 511 AA.
AC Q70371;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DE AgCP11563 (Fragment).
GN Name=agCG45469; ORFNames=ENSANG00000019755;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anophelinae.
OX NCBI_TaxID=180454;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: Contains 8 WD repeats.
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008964; EAA12470.1; -.
DR InterPro; IPR006594; LSH.
DR InterPro; IPR001680; WD40.
DR Pfam; PF00400; WD40; 8.
DR PRINTS; PR00320; GPROTEINRPT.
DR ProDom; PD000018; WD40; 3.
DR PROSITE; PS00896; LSH; 1.
DR PROSITE; PS00678; WD_REPEATS_1; 3.
DR PROSITE; PS00892; WD_REPEATS_2; 6.
DR PROSITE; PS0294; WD_REPEATS_REGION; 1.
KW Repeat; WD repeat.
FT NON TER 1
SQ SEQUENCE 511 AA; 55631 MW; 7C4C2670316A8580 CRC64;
Query Match 80.1%; Score 2937.5; DB 2; Length 511;
Best Local Similarity 75.6%; Pred. No. 4e-50;
Matches 405; Conservative 46; Mismatches 30; Indels 55; Gaps 24;

QY 1 MSISDEVNPLVRYLOESGFSHSAFTGIESHSOSNINGALVPPAALISIIKGLQYV 60
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 8 MSISDEVNPLVRYLOESGFSHSAFTGIESHSOSNINGALVPPAALISIIKGLQYV 67
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 EAEVSINEDGTLFDGRPIESLSLIDAMPDVVOTROQAVYDKLAQQA-AAA-AAA-AAA- 117
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 68 EAEISIGEDGTEQL--G---ESLSLIDAMPVEVNRQ-----NM---Q----- 104
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 118 TNOQSSA-KNGE--NTANG--BENGAIH--TIANNHT-----DMVEVDGVEIP-SNK 161
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 105 -NQKQATKT-EPPET-NGTTEE-PAPPPPTTTTATT-TPAPVETMEVDQSIETIPAS-K 158
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 162 AVTLRGHESEVFICAMNPVS-DLLVSGSGDSTARIWNLSNSTSGPTQLVLRHCIRREGQ 220
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 159 AVTLRGHESEVFICAMNP--STDLLASGSGDSTARIWMSDN--PANPNQLVLRHCIRREGQ 216
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 221 DVPSNKDVTSLDWNSECTLLATGSDYGFARITWTKDGNLASTLGQHKGPFIKWKKNKGN 279
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 217 EVPSNKDVTSLDWNSECTLLATGSDYGFARITWTKDGNLASTLGQHKGPFIKWKKNKGN 275
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 280 FILSAGVDKTTIWDATGCE-AKQPPHSPALPDVDWQSNNTFASCSTDMCIHVCKLGQ 338
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 276 YILSAGVDKTTIWDATGCT--QQPSFSPALPDVDWQSNQSFASCSTDDQCIHVCKLG 334
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 339 DRPIKTFQCHTEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNCVHDLQAHNKEIYTIK 398
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 335 DRPIKTFQCHTEVNAIKWDPTGNLLASCDDMTLKIWSMKQDNCVHDLQAHNKEIYTIK 394
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 399 WSPGTGCTHWNTOGALVHSYRGTTGGIPEVCWNAAGDKVGASASDGSVCVLDLRK 514
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 395 WSPGTGCTHWNTOGALVHSYRGTTGGIPEVCWNAAGDKVGASASDGSVCVLDLRK 510
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 9
EBI DROME
ID EBI DROME STANDARD; PRT; 700 AA.
AC Q95RJ9; Q9XZK1;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE F-box-like/WD-repeat protein ebi.
GN Name=ebi; ORFNames=CG4063;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.; FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP SPECIFICITY, AND MUTANT EBI4.
RX MEDLINE=99234084; PubMed=10215623;
RA Dong X., Tsuda L., Zavitk K.H., Lin M., Li S., Carthew R.W.,
RA Zipursky S.L.;
RT "ebi regulates epidermal growth factor receptor signaling pathways in
RT Drosophila.";
RL Genes Dev. 13:954-965 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
```

RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Anantides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harzisi M.,  
 RA Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong N., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).  
 RN [4]  
 RP FUNCTION, INTERACTIONS WITH PHYL AND SINA, AND MUTANTS EBIC1 AND  
 RP EBIC3.  
 RX PubMed=11032805;  
 RA Boulton S.J., Brook A., Staehling-Hampton K., Heitzler P., Dyson N.;  
 RT "A role for Ebi in neuronal cell cycle control.";  
 RL EMBO J. 19:5376-5386(2000).  
 RN [5]  
 RP FUNCTION, AND INTERACTIONS WITH SNO; SMR AND SU(H).  
 RX PubMed=12230979;  
 RA Teuda L., Nagaraj R., Zipursky S.L., Banerjee U.;  
 RT "An EGFR/Ebi/Sno pathway promotes delta expression by inactivating  
 RT Su(H)/SMRTER repression during inductive notch signaling.";  
 RL Cell 110:625-637(2002).  
 CC -1- FUNCTION: F-box-like component of E3 ubiquitin ligase complexes;  
 CC involved in R7 photoreceptor cell differentiation, cone cell  
 CC development and neuronal cell cycle control. E3 ubiquitin ligase  
 CC complexes mediate ubiquitination and subsequent proteasomal  
 CC degradation of target proteins. Required for specification of R7  
 CC photoreceptor cell fate in the eye by participating to the  
 CC ubiquitination and subsequent proteasomal degradation of trantrack  
 CC (ttk), a general inhibitor of photoreceptor differentiation.  
 CC Required to block the S phase entry in the peripheral nervous  
 CC system and central nervous system in a process that does not  
 CC involve the degradation of ttk. Involved in cone cell development

CC by preventing the transcriptional repression mediated by Su(H) on  
 CC D1, probably by participating to a E3 complex that contains sno  
 CC and mediates the ubiquitination and subsequent proteasomal  
 CC degradation of some component of the Su(H) repressor complex.  
 CC -1- SUBUNIT: Component of some E3 complex at least composed of sina,  
 CC ebi and phyl, required for the degradation of ttk. Probably forms  
 CC a E3 complex with sno, required for the degradation of some  
 CC component of the Su(H) repressor complex. Interacts with sno and  
 CC Su(H) and Smr.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- TISSUE SPECIFICITY: Widely expressed both in embryos and larvae.  
 CC -1- DOMAIN: The F-box-like domain is related to the F-box domain, and  
 CC apparently displays the same function as component of ubiquitin E3  
 CC ligase complexes.  
 CC -1- MISCELLANEOUS: 'Ebi' means 'shrimp' in Japanese.  
 CC -1- SIMILARITY: Belongs to the WD-repeat EBI family.  
 CC -1- SIMILARITY: Contains 1 F-box-like domain.  
 CC -1- SIMILARITY: Contains 1 Lish domain.  
 CC -1- SIMILARITY: Contains 8 WD repeats.  
 CC -1- CAUTION: Ref.3 sequence differs from that shown due to a  
 CC frameshift in position 484.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; AF146345; AAD35017.1; --  
 CC EMBL; AB003589; AAF51501.1; --  
 CC EMBL; AY061326; AAL28874.1; ALT\_FRAME.  
 CC HSSP; P16649; 1ERJ.  
 CC FlyBase; FBgn0023444; ebi.  
 CC GO; GO:0000074; P:regulation of cell cycle; IMP.  
 CC InterPro; IPR006594; Lish.  
 CC InterPro; IPR001680; WD40.  
 CC Pfam; PF00400; WD40; 3.  
 CC PRINTS; PR00320; GPROTEINRPT.  
 CC ProDom; PD000018; WD40; 3.  
 CC SMART; SM00667; Lish; 1.  
 CC SMART; SM00320; WD40; 8.  
 CC PROSITE; PS50896; LISH; 1.  
 CC PROSITE; PS00678; WD\_REPEATS\_1; 3.  
 CC PROSITE; PS50082; WD\_REPEATS\_2; 6.  
 CC PROSITE; PS0294; WD\_REPEATS\_REGION; 1.  
 CC Cell cycle; Nuclear protein; Repeat; Ub1 conjugation pathway; Vision;  
 KW WD repeat.  
 FT DOMAIN 4 36 Lish.  
 FT REPEAT 41 84 F-box-like.  
 FT REPEAT 353 392 WD 1.  
 FT REPEAT 408 447 WD 2.  
 FT REPEAT 449 488 WD 3.  
 FT REPEAT 491 531 WD 4.  
 FT REPEAT 532 571 WD 5.  
 FT REPEAT 574 622 WD 6.  
 FT REPEAT 625 664 WD 7.  
 FT REPEAT 666 700 WD 8.  
 FT DOMAIN 206 323 Ser-rich.  
 FT MUTAGEN 16 16 L->Q; In ebiC1; induces ectopic S phases  
 FT within the peripheral and central nervous  
 FT system.  
 FT C->Y; In ebi4; induces defects in R7 cell  
 FT development in the eye.  
 FT S->L; In ebiC3; induces ectopic S phases  
 FT within the peripheral and central nervous  
 FT system.  
 SQ SEQUENCE 700 AA; 72387 MW; 28C6D8D07BB79FB7 CRC64;  
 Query Match 79.5%; Score 2916.9; DB 1; Length 700;  
 Best Local Similarity 57.3%; Pred. No. 1.9e-49;  
 Matches 411; Conservative 51; Mismatches 34; Indels 221; Gaps 61;

QY 1 MSISSEVNFVLYRLOESGFSASFTFGIESHISQSNINGALVPPAALISIIQKGLQYV 60  
DB 1 MSFSSDEVNFVLYRLOESGFLSHAVVFGIESHISQSNINGALVPPAALITILQKGLYT 60  
QY 61 EAEVSNEDGTLPD-GRPIESLIDAVMPDV-----VOTRQO-----AYRD---KLA-- 104  
DB 61 EVMSVGEDG---EVARPIEGLSLIDAVMPDVPLKPIVKT-BFGKPGAV-DSSAP-AGG 114  
QY 105 -Q-----OH--A-A-A-A-A-----A-----AA--A--A--A 115  
DB 115 NQNNNAKPEIKIEPGTGVAGSAGNKIAGTGTGTPTDOSASEVDSSGNAANNAGGTVA 174  
QY 116 ---A---A-T---N-----Q-----QSAK--N-G-BN---T-- 130  
DB 175 GNGAGGAGQASTGCSNSTSTPAGGDLAAPCASQKQNSNEAGSSSSGAGNAGNATSTDD 234  
QY 131 -A-----NG-----EE-----N-----G-AH-----TI- 141  
DB 235 AASSTSTNGNSTSSSVEQPTSLGTAGGTVTSNPDAAASGAGSTATGSKAPSGAVTIR 294  
QY 142 --A--NN-----H-----T-DMMEV--DGD--VEIP-SNKAVALR 166  
DB 295 VQAQNNVQSGSSNAGSAPSGTISSTSGAGTTPAAL-VPMDIDENIEIPES-KARVLR 352  
QY 167 GHESEVIFICAWNVS-DLLVSGSGDSTARIWNLS-NSTSGPTQLVLRHCIREGGQDVPS 224  
DB 353 GHESEVIFICAWN-SDLLASGSGDSTARIWNDSANTNS--NQLVLRHCIOKGAEVPS 409  
QY 225 NKDVTSLDWNSEGTILATGYDGFARIW-TKQGNLASTLQHGKPIFALKWKK-GNFIL 282  
DB 410 NKDVTSLDWNCDGSLATGYDGYARIWKT-DGRLASTLQHGKPIFALKWN-KCNYIL 467  
QY 283 SAGVDKTTIWDATGE-AKQPFPHSAPALVDWQSNNTFASCSTD-MCIHVCKLG--Q 338  
DB 468 SAGVDKTTIWDASTQCT-QQAFHSAPALVDWQTNQAFASCSTDQR-IHVCRLGVNE 525  
QY 339 DRIKTFQGHTEVNAIKWDPCTGNLLASCDDMTLKIWSMKOD-NCVHDLQAHKEIYTI 397  
DB 526 --PIKTFKGHTEVNAIKWCPQQLASCDDMTLKIWSNDRCC-HDLQAHKEIYTI 582  
QY 398 KWSPTGFGTNNPNNMLASASFDSTVRLWDVDRGICHTLTKHQBPFVSVAFSPDGRYL 457  
DB 583 KWSPTGFGTNNPNTLILASASFDSTVRLWDVERGSCIHTLTKHTEPFVSVAFSPDGKHL 642  
QY 458 ASGSFDPKCVHINWOTGALVHSYRGTCGTFEVCNNAAGDKVGASDGSCVVLDIRK 514  
DB 643 ASGSFDPKCVHINWSTQGLVHSYKGTGTFEVCNWSKGTKVGSASDGSVFVLDIRK 699

RESULT 10  
Q96RW7  
ID Q96RW7 PRELIMINARY; PRT: 5636 AA.  
AC Q96RW7;  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Hemiscenins.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Trent J.;  
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF156100; AAK68690.1; -  
DR HSSP; P07996; 1LSL  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0005509; F:calcium ion binding; IEA.  
DR GO; GO:0005021; F:vascular endothelial growth factor receptor. .; IEA.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
DR InterPro; IPR000152; Asx\_hydroxyl\_5.

DR InterPro; IPR000875; Cecropin.  
DR InterPro; IPR000742; EGF 2.  
DR InterPro; IPR001881; EGF\_Ca.  
DR InterPro; IPR006209; EGF\_like.  
DR InterPro; IPR006605; G2F.  
DR InterPro; IPR009030; Grow\_fac\_recept.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003598; Ig\_c2.  
DR InterPro; IPR000169; Pept\_cys\_accsite.  
DR InterPro; IPR000884; TSP1.  
DR InterPro; IPR009134; VEGFR.  
DR InterPro; IPR002035; VWF\_A.  
DR Pfam; PF07645; EGF\_CA; 8.  
DR Pfam; PF07474; G2F; 1.  
DR Pfam; PF00047; IG; 44.  
DR Pfam; PF00090; TSP 1; 6.  
DR PRINTS; PR01832; VEGFRECEPTOR.  
DR SMART; SM00179; EGF\_CA; 7.  
DR SMART; SM00408; IGC2; 43.  
DR SMART; SM00209; TSP1; 6.  
DR SMART; SM00327; VWA; 1.  
DR PROSITE; PS00010; ASX\_HYDROXYL; 5.  
DR PROSITE; PS00268; CECROPIN; UNKNOWN\_1.  
DR PROSITE; PS01186; EGF 2; 3.  
DR PROSITE; PS00026; EGF 3; 5.  
DR PROSITE; PS01187; EGF\_CA; 8.  
DR PROSITE; PS050835; IG LIKE; 44.  
DR PROSITE; PS00639; THIOL\_PROTEASE\_HIS; UNKNOWN\_1.  
DR PROSITE; PS00092; TSP1; 6.  
KW EGF-like domain.  
SQ SEQUENCE 5636 AA; 613673 MW; F000B319CED7B52C CRC64;  
Query Match 71.6%; Score 2627.8; DB 2; Length 5636;  
Best Local Similarity 9.6%; Pred.No.1.3e-41;  
Matches 438; Conservative 59; Mismatches 15; Indels 4072; Gaps 372;  
QY 1 M-S-----I-----SS---D-----EV----- 8  
DB 1 MISWEVHTVFLFALLYSSLAQDASQSEIRAEFPPEGASTLAFVFDVTGSMYDDLQVI 60  
QY 9 -----NF-LV----- 12  
DB 61 EGASKILETSLRKPRLFPFDPDPDPDPDPDPDPDPDPDPDPDPDPDPDPDPDPDPDP 120  
QY 13 -----Y-----R-----Y-----LO----- 17  
DB 121 MSIGAIIKALEISLPGSFYVFTDARKSKDYRLTHEVLIIQKQSQVVFVLTGDCDDRTH 180  
QY 18 -----E-----SG--F-----S--H--S-----A-----F 26  
DB 181 IGVKYVEEIASTSSGQVFLDKKQVNEVLKWVEEAQASKVHLLSTDHLEQAVNTWRIPF 240  
QY 27 -----T-----T-----FG-----I-----E-----B----- 31  
DB 241 DPSLKEVTVSLSGSPSPMIBIRNPLGKLIKKGFGELHELLHNSAKVNVNKEPEAGMWTVK 300  
QY 32 ---S--H-----I-----S--Q----- 36  
DB 301 TSSSGRHSVRITGLSTIDFRAGRSRKPTLDFKKTVSRRPQGIPTYYVLLNLTSGISTPARID 360  
QY 37 -----S-----S-----NI-----N-----G----- 41  
DB 361 LLELLISGSSSLKTPVKYYPHRKPYGIWNIISDFVPPNEAFFLKVTGYDKDYLFQVSS 420  
QY 42 -----A-----L-----VP-----P-----A 47  
DB 421 VSFSSIVDPAPKVTWPEKTPGYVLPQGPICSDVSLPFTLSFVRNGVTGLGVQYLKESA 480  
QY 48 -----A--L-----I-----S-----IIQ-----KG----- 56  
DB 481 SVSLDIKAVTLSDGEFYECIAVSSAGTGRAQTFDFVSEPPPIQVPPNVTVTGERAVLT 540  
QY 57 -----L--Q-----Y--V--E----- 61



Db 541 CLIIAIVDYNLTQRNDVRVLAEPARIRTLANLSLELKSVDKENDAGEYHCWVSSEGGSS 600  
QY 62 -A-----EVS-----NE-----68  
Db 601 AASVFLTVQBPVKVTVMPKNQSFSGSEVIMSATGYPKPKIAWTVNDMFIVSHRYM 660  
QY 69 --DGLP-----G-----R-----P-----I--78  
Db 661 TSDGTLFIKNAAPKADAGIYCLASNSAGTDKQNSTLYRIEAPKLMVQVSELLVALGDITV 720  
QY 79 -E--S-----L--S--LI-----DA-----V-----87  
Db 721 MECKTSGIPPPQKWKFGDLELSPSTLIDPLGLLKIQTQDLDAAGDYTCVAINEAGR 780  
QY 88 -----M-----P--DV-----VQ-----T-----94  
Db 781 ATGKITLDVGSPPVFIQEPADVSMEIGSNVTLPCTYVQGYPEPTIKWRRLDNMFI SRPFS 840  
QY 95 -----R-----OH-----A-----108  
Db 841 VSSISQRTGALFILNLWASDKGTIYCEAENQFGKIQSETTVTVTGLVAPLIGISPSVAN 900  
QY 96 ---QQ-----A-----Y--R--D--KL-----AQ-----105  
Db 901 VIEGQQLTLPCTLLAGNPIPERWIKNSAMLLQNPYITVRSDSLHIERVOLQDGGEYTC 960  
QY 106 -----OH-----A-----108  
Db 961 VASNVAGTNNKTTSVVVHVLPTIHQGQILSTIEGIPVILPCKASGNPKPSVIWSKKGEL 1020  
QY 109 -----A--A--A-----A--A--A-----114  
Db 1021 ISTSSAKFSAGAGSLYVSPGEESGEYVCTANTAGYAKRVQLTVYVPRVFGDLRG 1080  
QY 115 -----A-----A-----116  
Db 1081 LSQDKPVEISVLAGEEVLPCVKSLPPPIITWAKETQLISPFSPRHTFLPSSGSMKITET 1140  
QY 117 -----ATN-----Q-----Q--GA--K-----N-----126  
Db 1141 RTSDSGMYLCVATNIAGNVQAVKLVNHVPPKIQRGPKHLKVQVGORVDIPCAQAQTPLP 1200  
QY 127 -----GE--N-----T--A--N--G-----133  
Db 1201 VITWSKGGSTMLVDGEHHVNDPDTLSIDQATPSDAGIYTCVATNIAGTDETEITLHVQE 1260  
QY 134 -----E--N-----GA-----H-----139  
Db 1261 PPTVEDLEPPYNTTFQERVANQRIEPPCPAKGTPKPTIKWLHNGREL TGREP GISILEEG 1320  
QY 140 T--IA-----N-----N--HT-----DM-----148  
Db 1321 TLLVIASVTPYDNGEYICVAVNEAGTTERKYNLKVHVPVVIKQEQVNSVLLNQLTNL 1380  
QY 149 -----M-----E-----150  
Db 1381 FCEVEGTPSPIIMWKNDVQVTSSTIQTVNNKILKLFATPEDAGRYSCKAINIAGTS 1440  
QY 151 -----VD-----G-----DV--E-----I-----P--158  
Db 1441 QKYFNIDVLPPTIIGTNFPKEVSVLNRDVALEQVKGTFPPDIHWFKDGKPLFLGDPN 1500  
QY 159 -----V-----SN--K--A-----V-----VL 165  
Db 1501 VELLDRGQVLHKNARNDKGRYQCTVSAAGKQAKDIKLTIVIPPSIKGNVTTDISVL 1560  
QY 166 -----RG-----H-----168  
Db 1561 INSLIKLECTRGLPMPAITWYDQGPIMSSQALYIDKQOYLHI PRAQVSDSATYTCVH 1620  
QY 169 -----E--S-----E-----VFI-----C--A-----W-----177

Db 1621 ANVAGTAERKSFHVVDVYVPPMIEGNLATPLNKQVVIASHSLTECNAAAGNPSPIITLWKDGV 1680  
QY 178 -----N-----P--V--SD--182  
Db 1681 PVKANDNFRIAGGKKLEIMSQAQIEDRGQYICVATSVAGEKEIKYEVDVLVPPAIEGGDE 1740  
QY 183 -----LL-----VSGS-----G-----D-----S--191  
Db 1741 TSFYIWMNNLLELDCHVTGSPPTIMLWKDGLIDERDGFKILLNGRKLVIQAQVSN 1800  
QY 192 -----TA-----R-----I--W 196  
Db 1801 GLYRCMAANTAGDHKKEFEVTVHPPTIKSSGLSERVVVKYKVALQCIANGIPNSITW 1860  
QY 197 -----NL-----S-----E-----NS--T-----S 204  
Db 1861 LKDDQVNTAQGNLKTQSSGRVLQIAKLTLEDAGRYTCVATNAAGETQQHQLHVHPPS 1920  
QY 205 -----G-----PTQL-----V-----L--R-----212  
Db 1921 LEDAGKMLNETVLVSNPVQLECKAAGNPVPIWYKDNCLLSGSTSTMTFLNRQIIDIES 1980  
QY 213 -----H-----C-----I--215  
Db 1981 AQISDAGIYKCAINSAGATELFSYLSQVHVAPISGSNNMVAVVWNPVRLECEARGIPA 2040  
QY 216 -----R-----E--G-----G--Q--D--221  
Db 2041 PSUTLWKDGPVSSFNGVLQVLSGGRIALATSTQISDTGRYTCVAVNAAGEKORDILRV 2100  
QY 222 -VPSN-----KD-----VT-----S--L--231  
Db 2101 YPPNMGBEQNVSVLISOAVELLQSDAIPTTLTWLKDGHLLKPKGLSISENSVLK 2160  
QY 232 -----D-----W--N--S-----EG--TLL--240  
Db 2161 IEDAQVQDGRYTCVATNVAGTEKKNVNVNVPNIIGSGDELTQLTVEGNLISLCE 2220  
QY 241 AT-----GS--Y--D-----G-----F-----A--249  
Db 2221 SSGIPPNLIMKKGSPVLTDMSGRVRIILSGGRLQISIAEKSDAALYSCVANSVAGTAK 2280  
QY 250 -----R-----I-----WTKDG-----256  
Db 2281 KEYNLQVYIRPTITNSGSHPTIIVTRGKSISLECEVQGPPTVTWMDGHPILIRAKGV 2340  
QY 257 -----NL--AS-----T-----L-----GOH-----265  
Db 2341 EILDEGHILQLKNIHVSDTGRYVCVAVNVAGMTDKKYDLSVHAPPSIIGNHRSPENISVV 2400  
QY 266 -K-----G--PI-----FALK--W-----274  
Db 2401 ERNSVLTCEASGIPLPSTTWFF--KDGWPVLSVNSVRILSGGRMLRMQTTMEDAGQYTC 2458  
QY 275 -----N-----K--K-----GN-----279  
Db 2459 VVENAAGEERKIFGLSVLPVPHIVGENTLEDVKVKEQSVTLTCEVTGNPVPETWHKOG 2518  
QY 280 -----F--I-----L--S--AGV--286  
Db 2519 QPLQDEAHIIISGGHFLQITNVQVPHTRGYTCCLASSPAGHKSRFSLSLVFVSPTIAGVG 2578  
QY 287 -D--K--T-----T--I-----I--W-----293  
Db 2579 SGNPEDVTILNSPTSLVCEAYSYPATITWFKDGTPLSNRNRILPGRTLIQLINAQ 2638  
QY 294 -D-----A-----H-----TG-----E 299  
Db 2639 EDNAGRYSCVATNEAGEMIKHYEVYIPIINKGDLWGPGLSPKEVKVNNVTLTLECE 2698  
QY 300 A-----K-----Q--Q-----303  
Db 2699 AVAIPASLSWYKDGOPKSDDDHVNIAANGHTLIQIKEAQISDTGRYTCVANSIAGEDELD 2758







Db 10403 DPENPHRLAPTALGEVIOGPTILREYLSVDORTEAAVTKSLPDWAPFREQSGSRFNK 10462  
Qy 162 -----AV-V-L----- 165  
Db 10463 SGDLGVNPDGTIEFSSKDTQVKIRGLRVELGEVHAHQVALDGVHQIAVDVFKDNGT 10522  
Qy 166 --RG--HE--S--E--V-----FI-C-- 175  
Db 10523 QPRGYFSDESROIHEAHPGSPQAIDKQLARLTVAIGKLNIALPRYMIPTLPCKY 10582  
Qy 176 -----A-----W-----NP-- 179  
Db 10583 MPSITSTKDRNELRRRTLLTQSELAMFYSRCNKRTPEPMESLIQRIWSEILHPLDS 10642  
Qy 180 -----VS-----D-L-----L----- 184  
Db 10643 IGRDSDFLGGSITAIHLGSISSROGISTAKOVDFDPLNDIASKARELDEAQOQS 10702  
Qy 185 --VS--GS-----G----- 189  
Db 10703 PDISPFSLIGETRELAAGSVKKQLKLTSTQIEDAYPCTSKI PGSYIAKYVYRLPHOV 10762  
Qy 190 D-----S--T-A-----RI-----W-NL-S-----E-N-----S 202  
Db 10763 DIGRFKSSMEHTVALHPILRLTRIIMIKOTCVQLVVKDSVTWENLDSGDLDEAVNATHSYS 10822  
Qy 203 -T-----SG-----P----- 206  
Db 10823 MTGSPLSRYSIHESKSGDYFLWTAHSHIDGWSVPVINTLYQAYKGLGPKAYSG 10882  
Qy 207 -T-----OL----- 209  
Db 10883 FIKYTMEHOKAAGDYWEQLQNAKRASFPKTKLAAKSIWTRKTRVMTSLKFSASSNE 10942  
Qy 210 -----H-----V-L-----R----- 212  
Db 10943 IATKASVIRAAMAVILARYCESDDVCFGATISGRQASVPGLEMGAPAVATVPRVQLDN 11002  
Qy 213 -----H-----C-----I-R----- 216  
Db 11003 DQEISKFLQNIQSOAHMPYEQYGLQSTAKLGNARDACDFTSLVVIQPIQOEFSVOS 11062  
Qy 217 -----E-GG-----Q--D-----V-----P-- 223  
Db 11123 IVOQLLAEDGGKLGDISIASSPDLBAQKSNQDGPQIIDDCIHLIIERQAKQRPNRPAT 11182  
Qy 224 -----S-----NK----- 226  
Db 11183 DAWDGFTYSELDRNTANRLAHLVHDYAVKVGDIHVCFEKSRYVAILAVNKAGAANA 11242  
Qy 227 --D--V-----T-----S-LD--W--NSE- 236  
Db 11243 PEDPAHPQRLKAVASQTKAKLALASTANTRLCEQVQVDCVVESSFLDKNLWTYDIEK 11302  
Qy 237 G-----L-----L-----A-----T----- 242  
Db 11303 GPDINVTMDAAVILFTSGSTGVPGKIVMGHGCALCTNOAALSGWLGFHTVRMLQPSFV 11362  
Qy 243 -----G-----S-----Y-----D----- 246  
Db 11363 FVSVGEIWOALMNGACVCPSEHMLNSLDSFVRDFNTWYALTPTSFTTLKPKDFPSL 11422  
Qy 247 -----G-----F-----A-----R----- 250  
Db 11423 KULLLAGEPTQDVLDTWGLPNTRFINANGPAETCVYNTLYEWQNTSPKLGRAVCA 11482  
Qy 251 -IW----- 252  
Db 11483 YIWWVDVENPQRLAPTGLGEIIVQGPPLLKEYLADPEKTAATAATVTELPENAPRROSTTW 11542

Qy 253 -----T-----KD----- 255  
Db 11543 NRYRTDGLGFDYDHDGMLHFASRKDTQVKIRGLRVELGEVEHRIQSSLDGVRQVAVDVPK 11602  
Qy 256 --G-NL--AS--T-----LG-- 263  
Db 11603 TEKANLVAYLCTSDTKTPDQNTDFEGKDVFASIDTELQINLGKMLTELNSSLPTYMIP 11662  
Qy 264 -----Q----- 264  
Db 11663 TMFPCAYMPFITSSKLDVRKRLRLTAELSQDOLEAYSLDVNKKQAPETEMEIRLOKLWA 11722  
Qy 265 -----H-----K--G----- 267  
Db 11723 EVLDLPEASIGRHDNFRIGGDSIAAIRLVSMARDAGISLTVNDIPDDARLISVATKAID 11782  
Qy 268 -----PI-----FAL----- 272  
Db 11783 NDESHLMAPIEPFSLTSGIEDFALPSLSDQFIEDAYPCSKLOEGLMALAIKOPGSY 11842  
Qy 273 -----K-----W-----N----- 275  
Db 11843 IAKYIYKLSEHVDVDFKAAWESTVEIASALRTRIIQTGCSIQVINGDISWDEADGDL 11902  
Qy 276 -----KKG-----N--F-----I--LS-----AGV 286  
Db 11903 QSCLAQSTLEMGYTRLCRYTLIRRDNDTYFLWNMHHAVIDGLSTQNLGTLFNIYSGV 11962  
Qy 287 -----DK----- 288  
Db 11963 DVLPLPPNRFQYMLQLEEDTANTYWKQLHNAQRTIPPNVDVARDKPAATQMLQSSIE 12022  
Qy 289 -----T-TI--I--W----- 293  
Db 12023 LPHGLDNSTVTIATVIRAAWAILARYCSDDDVTFTTISGRQAPIEIMGVGTIATV 12082  
Qy 294 -----DA----- 295  
Db 12083 PVRVLNROQLISDFLEGVQRAVEMIAFEQYGLQNIAKLGDDARDACDFSSLLVQPIQ 12142  
Qy 296 H-----T-----GE--A--K--Q-----Q-----F----- 304  
Db 12143 HLAGTKDSDILVDASIGEDNAVEAIQNYFSYPLVIAQHLHYDDHIFILYDSSIMLEPQ 12202  
Qy 305 -----P-----F----- 306  
Db 12203 MVALSYHLOHVMQLTNTTTPATLETVSVSSYDVERALAINABEVIDTCMHMFENRV 12262  
Qy 307 H-S--APA-----L-D--V--D----- 315  
Db 12263 HLNPLAPAAWDGNTYAEALNSAANKLAHYLIKSYSVKLNLDLVHVCFKSAWYIVSIIA 12322  
Qy 316 -----W-----Q-----S--N-----N----- 320  
Db 12323 INKAGATVPLDPFPHEQELRSIVNQTKTALASPGNITLCSALVDNVVVESSALDSML 12382  
Qy 321 -----T-----PAS-----C----- 325  
Db 12383 PAQDGLSPKSVSSRTAAIYILFTSGSTGTPKGVIOHRSCTNMMNAIAKRVRYHLDVRI 12442  
Qy 326 --ST--D--D-----M-----C----- 330  
Db 12443 LOFSAYVDFDSILEIIMSLLOGACVCPSEHIRMNGIVDFIRDMINMLYLTFSFLRTIN 12502  
Qy 331 -----IH----- 332  
Db 12503 PIDVNVNVLVAGGEAI PRDVFETWGVRFVINGWGTPTTETTVGSIHEFESVDESPTIG 12562  
Qy 333 --V--C-----K--LG-----Q-----D-----RP 341  
Db 12563 HPVGGFCMIVDPNNPQLLAPTGLGEIIVQGPPLLHEYLNDNPKTQEAAILYDLPEWAPRP 12622

QY 342 ---IKT---F---Q--- 346  
Db 12623 DENMGRFYKTGDLGFYNANGKIEFSRKDTQVKIRGLRVELGEIEYQVQASVEEIRQIA 12682  
QY 347 ---G---H---TN--E---V--- 352  
Db 12683 VDVIKTDGNSLVAYLCFNDEIRQLNADVNGPFPPLDNLQETLAGAIGKLSVTLPRYM 12742  
QY 353 ---N---A--- 355  
Db 12743 IPTFYPCSYMPSITSGKLDRLKELKQTAALSQSELNKFSLHGVSKRAPETPMELQONI 12802  
QY 356 KW---DP--- 359  
Db 12803 -WSKLASIPESIGRDDSLFLGGDSIMAHVLTACREAGVSLTVKEIIFDDPRLSAVASH 12861  
QY 360 ---T--- 360  
Db 12862 ARSMDAIDQDLPLVPLSLLSDRLREMLVSDDTFQTSLLPSQVIEDAYPCSKLOEGLMAL 12921  
QY 361 ---G---N---LL--- 364  
Db 12922 SVKQGSYVAQYVYKSSGVDLVRFKASWERTIQLCANLRTRIVMLDGTGVLLVDGPAB 12981  
QY 365 ---A-S---C---S---D---M- 371  
Db 12982 WDNAFSTDQCLTNATREMYGSLRNRYALVRDPEFGNHFIVSHHAVHDGWTLRIMN 13041  
QY 372 TL--- 373  
Db 13042 TLYSLYDQGLPMLPYSAFIRYTVNIDKEASRFWTEQMKNAKRATYPPMPRIESHTGV 13101  
QY 374 ---K-I---W---S---M 378  
Db 13102 SRMMNMTISPPSVKTVNKTALRGWAILLARYCDTDDITFGTTVSGRQAPVPGITEM 13161  
QY 379 ---K---Q-D---N---C--- 383  
Db 13162 PGVAVTVPIRVLEASQTVQAFSLNQSQATDMIAPEQFGLQNIKLTSDARDACEFSS 13221  
QY 384 -V---H-DL--- 387  
Db 13222 LLVQPRSHLDLLKKGESSESLVASVARSQGLQNYFTYPIVQGHVFEDSIELLLT 13281  
QY 388 ---Q-A-H---N---KE--- 393  
Db 13282 YDSTILSEVQMKALSHQFNVANQLVNESNDPLSSITISGEWDEQAKWNVENPEILDT 13341  
QY 394 ---I---YT---IK--- 398  
Db 13342 CIHSLIEQAIRPDAPAI CANGEMNYSOLNSAANQLAHHILKIGIKADDLVHVCPEKS 13401  
QY 399 -W---S---PT---G---PGTN---N-- 408  
Db 13402 WNFVSVIIANKYGAAMVPLDPSHPEQRLQVQVQTLAKFALSSP-TNAALCNKLNVHVI 13460  
QY 409 ---P---N-A---N---L-M---LA 416  
Db 13461 EVSPSLIDELSKFCDFGSPAINVPSSNAAVYVLTSGTGTGPKLVMOHGAVCTSOATA 13520  
QY 417 ---S---AS-FD---S-T---V--RL-W- 427  
Db 13520 KRSLTDPVRILOQAAVYFVDSIGEIVAPLIHGACVCPSEETRMNGLKEFIRDARINWA 13580  
QY 428 ---DV--- 429  
Db 13581 YLTPSFVRLTPEDVPSIQLLLLAGAVGRDILDTWFGKVLINGWGPACTCVSTLHEW 13640  
QY 430 ---D---R-G---I-C--- 434  
Db 13641 SSIDESPLTIGRPVGGCWIVEABDSNKLTPIGCLGEVVLQGPFTLLREYLADPQSKETI 13700  
QY 435 I-----H-----TL-----T-K----- 440

Db 13701 ITELPPWAPQVSDAHLWSRFYKSGDLCFYNPENGTLFFYSRKDTQVKIRGLRVELGEVEH 13760  
QY 441 H---Q---E---P---VY--- 446  
Db 13761 HIRELLEGVRQVAVDLTSETGTQLVSVYICFNDDSQSPSPELKASDIYLPDLADIOARIT 13820  
QY 447 ---SVA---F--- 450  
Db 13821 SMVGELSVTLPRYMTPTLPIPKYMPFTTSTKLDKRLTKLSITASLGRDELIHYLSLNSKK 13880  
QY 451 ---S---P-D-GR---YLAS---G-S--- 461  
Db 13881 RVPETEMETRLQAIWSEILNLPIDISIGRDDSLFQIGGDSITAIYLVSKAREAGISLIKD 13940  
QY 462 -FD---K-C-V---HI--- 468  
Db 13941 IFDDSRLLAVASKAVLSTEFQBPQGPVFPFSLNKLQTRALVAGEVRKFCGLNDDHIED 14000  
QY 469 ---WN--- 470  
Db 14001 AYPCTSLQELMALTWKQPGSVYAKYVYKLSFTVDMERFOAAWNRMTMELCGNMRTRIVLL 14060  
QY 471 --T--Q---T---GA---L---V--- 477  
Db 14061 NGTPIQLLLKEDNQWOLENDTLASITNSRDLKMGYGAPLCWYGVLEENNNKYFVWSAH 14120  
QY 478 HS---Y--- 480  
Db 14121 HSYDGMVNRILLTTLTYITLSTVPTLQPSGFIKYNMELDSMALADFWRELQSGSKRA 14180  
QY 481 ---R--- 483  
Db 14181 VFPAROTSPSSSSTQIFKSIISLEQAKOSIITKASILRAAWAIVLARYCDTDDVSGFTT 14240  
QY 484 ---G--- 484  
Db 14241 VSGRHAPVAGLETMPGPMIATVPVRVHLNRSATKSQFLTEIQKQAYEMVPEYQGLQNIS 14300  
QY 485 ---G--- 486  
Db 14301 KLSRDARTCDFSSLVIOPPATTISEKADNTNIIYDEBOSLTDAMHINYFNYPLVII 14360  
QY 487 ---FS--- 489  
Db 14361 LMTFEDHILORFNFEVLTARVSALSNNHIDHVVKLLAETDETLESVNLVGDWDIOHA 14420  
QY 490 ---C-W---N-AA---GD---K---VG 499  
Db 14421 LASTRLKPSTESCTHMQIERIKTQPNQDIAIVSWDGLTYKELGVLASRLAWKLGIVG 14480  
QY 500 ---AS--- 501  
Db 14481 PESLIPLCFPKSTWAVVAVMAIEMAGGAFVPLDPNAPVARLGIIEBTKSSSLAVASPSQC 14540  
QY 502 ---A--SD---GS---V--- 507  
Db 14541 DTWRDIGIEVFAVDEALLLELSDPADGVKSMQAPDNASVVLFTSGTGKPKGMVQHNSL 14600  
QY 508 ---C---VLD---L---R--- 513  
Db 14601 CSSGNAYGCDLEIGPQTRIFQFSAYTFDVGVLDCVLSMRGATICIPSDHARLNDLAGAM 14660  
QY 514 ---K 514  
Db 14661 NVTK 14664

RESULT 12  
AAP49011  
ID AAP49011 PRELIMINARY; PRT; 7073 AA.  
AC AAP49011;  
DT 02-MAR-2004 (TriEMBLrel. 27, Created)

DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DE 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
OS Orlab polyprotein.  
OC SARS coronavirus 2301.  
OC Viruses; serNA positive-strand viruses, no DNA stage; Nidovirales;  
OC Coronaviridae; Coronavirus; Human coronavirus (strain SARS).  
OX NCBI\_TaxID=230471;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ZJ01b;  
RX MEDLINE=22889812; PubMed=14527350;  
RA Li L., Wang Z., Lu Y., Bao Q., Chen S., Wu N., Cheng S., Weng J.,  
RA Zhang Y., Yan J., Mei L., Wang X., Zhu X., Yu Y., Zhang M., Li M.,  
RA Yao J., Lu Q., Yao P., Bo X., Wo J., Wang S., Hu S.;  
RT "Severe acute respiratory syndrome-associated coronavirus genotype and  
RT its characterization.";  
RL Chin. Med. J. 116:1288-1292(2003).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ZJ01b;  
RX PubMed=14733771;  
RA Wang Z.G., Li L.J., Luo Y., Zhang J.Y., Wang M.Y., Cheng S.Y.,  
RA Zhang Y.J., Wang X.M., Lu Y.Y., Wu N.P., Mei L.L., Wang Z.X.;  
RT "Molecular biological analysis of genotyping and phylogeny of severe  
RT acute respiratory syndrome associated coronavirus.";  
RL Chin. Med. J. 117:42-48(2004).  
DR EMBL: AY286320; AAP49011.4; --  
KW Polyprotein.  
SQ SEQUENCE 7073 AA; 790251 MW; 96E95991DACAGEA9 CRC64;  
  
Query Match 71.4%; Score 2620.7; DB 2; Length 7073;  
Best Local Similarity 8.3%; Pred. No. 2.7e-41;  
Matches 469; Conservative 34; Mismatches 6; Indels 5163; Gaps 389;  
  
QY 1 M-----S-----S-----DEV-----N----- 9  
DB 297 MGRIRSVYPVAPQECNNMHLSTMKNCHDEVSQTCDFLKATCSHCCTENLVIEGPTT 356  
  
QY 10 -----F-LV-- 12  
DB 357 CGVLPNAVVMPCAPQDEIGPEHSVADYHHNSHNIETRLKGGTRCFGGCVFAYVC 416  
  
QY 13 Y-R-Y-----L-Q--E-S-----G-FSH----- 23  
DB 417 YNKRAYVPRASADIGSGHTIGTGDVETLNEDLLEILSRVNIIVGDF-HLNEEVAI 475  
  
QY 24 -----SAP--T-----F-----G---I----- 30  
DB 476 ILASFSASTSAFIDTILSKLDYKSFKTIVESCNGYKVTGKPKVGAWNIQQRSVLPLCG 535  
  
QY 31 ---E-----S-----H-----IS-QS-----N-- 38  
DB 536 FPSQAAGVIRSIPTARDANHSIPDLQRAAVTILDGISEQSLRLVDAMVYTSDDLTVNSV 595  
  
QY 39 -I-----N--GALVP--P-----A-----A-----LIS- 51  
DB 596 IIMAYVTGGLVQTSQWLSNLLGTITVEKLKRPFEWIEAKLSAGVEFLKDAWEILKFLITG 655  
  
QY 52 -I-----IQ-----K-----G-----L-----Q----- 58  
DB 656 VFDIVRGQIQVADSNIKDCVKCFIDVNVNKALEMICDQVITAGAKLSRLNLSGEVFIQSKG 715  
  
QY 59 -Y-----V-----V-----EA-E--V-S-- 65  
DB 716 LYRQCIRGKEQLQLMLPLKAPKEVTFLEGSDHTVLTSEEVVLKNGELEALETPVDSFTN 775  
  
QY 66 -----I-----N-----ED----- 69  
DB 776 GAIVGTFVCVNGLMLEIKDKEQYCALSPCLLATNVNFRLLKGGAPIKGVTFGEDTVWEVQ 835  
  
QY 70 -----GT----- 71  
DB 836 GYKNVRITFELDERVDKLVNEKCSVTVESGTEVTFACVBAVAVTKLPVSDLLTNMG 895  
  
QY 72 -----LFD--G-----R-----P-----I----- 78  
DB 896 IDLDWSVATFVFLFDDAGEENFSRRMYCSFYPPDBEEEDAECEBEIDETCBHEYGTED 955  
  
QY 79 -----E--S-----L-----S-----L----- 83  
DB 956 DYQGLPLEFGASAEVVRVEEEDWLDTTTEQSEIEPBPPTPEBPVQFTGYLKLTDN 1015  
  
QY 84 --I--D-----A-----V-----M----- 88  
DB 1016 VAIKCVDIVKEAQSANPMVIVNANHLKHGGVAGALNKAATNGAMQKESDDYIKLNGPL 1075  
  
QY 89 -----P-----D----- 90  
DB 1076 TVGGSCLLSGHNLAKCLHVGPNLNAGBDIQLLKAAYENFNSQDILLAPLLSAGIFGAK 1135  
  
QY 91 -----V-VOT-R-Q-----QA--Y-----R----- 100  
DB 1136 PLQLSQVCVQTVRTQVYIAVNDKALYEQVMVDYLDNLKPRVEAPKQEPPTEDSKTEK 1195  
  
QY 101 -----D-----D-----KL--A-----O----- 105  
DB 1196 SVQKPDVVKPKACIDEVTTTLEETKFLTNKLLLFADINGKLYHDSQNLGDEMSFL 1255  
  
QY 106 -----O----- 106  
DB 1256 EKDAVYMGDVITSGDITCVVIPSCKAGGTTEMLSRALKKVPVDEYITTYPGGCGAGYTL 1315  
  
QY 107 -----HA-----A----- 109  
DB 1316 EBAKTALKKCSAFYVLPSEAPNAKEEILGTVSNLREMLAHAEETRLKMLPICMDVRAIM 1375  
  
QY 110 A-----A----- 111  
DB 1376 ATQRYKIKIKIOEGIVYGVRRFFFTYSKEPVASITKLSNLEPLVTPIGVYTHGFNL 1435  
  
QY 112 --AA-----A-----A-----A----- 117  
DB 1436 EEAARCSRSLKAPAVSVSSPDVATVYNGYLTSSSTKSEEHFVETVSLAGSYRDSVSGQ 1495  
  
QY 118 -----TN 119  
DB 1496 RTELGVFEFLKRGDKIVYHTLESPEFHLDEVLSDKLKLSLSREVTKIKVFTTVDNTN 1555  
  
QY 120 -----QQ-G-S--A-----K--N--G----- 127  
DB 1556 LHTQLVDMSTYGOQPGPTLDGADVTKIKPHVNHGKTFVFLPSDDTLRSEAFYHTL 1615  
  
QY 128 -E-----N-----T----- 130  
DB 1616 DESFLGRYSALNHTKKKPPQVGGILTSIKWADNNCYLSSVLLALQOLEVKFNAPALQEA 1675  
  
QY 131 -----AN-----GE-----E-----N--GA-----H-- 139  
DB 1676 YRARAGDAANFCAILAYSNKTVGELGVRETMTLHLOHANLESARVLNVVCKHCGQK 1735  
  
QY 140 -----T-----I----- 141  
DB 1736 TTTLTGVEAVMYGTLSDYDNLKTGVSIPCVGGRDATQYLVQESSFVMSAPPAEYKLOQ 1795  
  
QY 142 -----AN-----N-----H-----T-----D-----MM----- 149  
DB 1796 GTFLCANETGYVQCHYTHITAKETLYRIDGAHLTKMSEYKGPVTDVFKETSYTTIK 1855  
  
QY 150 -----EV-----DG-----D-----V8--I-----P-----SN----- 160  
DB 1856 PVSXKLDGVYTIETIEPKLDGYVKDKNAYVTEQPIDLVPTQPLPNASFDNFKLCSNTKFA 1915  
  
QY 161 -----K-A-----VV----- 164  
DB 1916 DDLNQMGTGFKPASRELSVTFPPDLNGDVVAIDYRHYSASFKKGAKLLHKPIVMHINQAT 1975

165 QY -----LR----- 166  
1976 Db TKTTFKENTWCLRLMSTKFPDTSNSPEVLAVEDTQGMNDLACESQOQTSEEVENPTIQ 2035  
167 QY -----GHE-----S-----E----- 171  
2036 Db KEVIEDVKTTVEVGNVILKPSDEGVKVTQELGHEDLMAAYVENTSITIKKPNLSLALG 2095  
172 QY -----N-----VF----- 173  
2096 Db LKTIATHGIAAINSVPWSKILAVYKPLQQAATTNSCAKRLAQRVNNMVPVFTLLFQ 2155  
174 QY -----I-----C-----A-W----- 177  
2156 Db LCTFTKSTNSRIRASLPTTIKNSKVSVAKCLDAGINYKSPKSKLFTIANWLLLSI 2215  
178 QY -----N-----P-----V-----S-D-L 183  
2216 Db CLGSLICVTAAGVLLSNFGAPSYCNGVRELYLNSNVTTMDPCEGSPFPCSICLSGLDSL 2275  
184 QY -----L-----V-----S-----G-S----- 189  
2276 Db DSYPALETIOVTISSYKLDLTILGLAEWVAYMLFTKFFYLGLSAIMQVFFGYFASHF 2335  
190 QY -----D-----ST-----A 193  
2336 Db ISNSLWMTIISIQMAPVSAMVRMYIFFASFYINKSYVHINDGCTSSITCMCMYKENRA 2395  
194 QY -R-----I-----WN-LSN-ST--SG----- 205  
2396 Db TRVECTITVNGMKRSFYVYANGRGFCCKTHNNCL--NCDTFCGTGTFISDEVARDLSLQ 2453  
206 QY -----PT-Q--L-----V-----L-----RH----- 213  
2454 Db FKRPINPTDOSSXIVDSVAVKNGALHLYFDKAGOKTYERHPLSHFVNLDNLRANNTKGS 2513  
214 QY -----C--I----- 215  
2514 Db PINVIVFDGSKCDESASKASVYYSQLMCOPILLDDQALVSDVGDSTEVSKWFDAYVD 2573  
216 QY -----R----- 216  
2574 Db TFSATFSPMEKLKALVATAHSELAKGVALDGLVSTFPVSAARQGVVDTDVTDKDVIECLK 2633  
217 QY -----E--G-----G-----O----- 220  
2634 Db LSHHSDLEVTGDCNNFMPLYNKVENMTPRDLGACIDCNARHINAQVAKSHNVSLIWNVK 2693  
221 QY ----- 221  
2694 Db DYNLSLSQLRKQIRSAKKNNIPRLTCATTQVNVNVTTKISKGGKIYSTCFKMLKA 2753  
222 QY -----V-P-----S-----N-----K--D-VT-----S-----LD- 232  
2754 Db TLLCVLAALVCYIWMVPVHTLSIHDGYTNEIGYKAIQDGVTRDIISTDDCFANKHAGFDA 2813  
233 QY -----N--S-----E-----GT----- 238  
2814 Db WFSQSGSYKNDKSCPVAAIITREIGFIVPGLPGVTLRAINGDFLHFLPRVFSAVGNIC 2873  
239 QY -----L-----LA-----T-----GS-----Y-- 245  
2874 Db YTPSKLIEYSDFATSACVLAECTIPKDMGKPVPCYDNTLLEGSISYSSELPDTRYVL 2933  
246 QY -DG-----F-A-----R--I-----W----- 252  
2934 Db MDGSIQFPNTYLEGSRVVVTTFDABYCRHGTCESEVGICLSTSGRWLNNEHRYALSG 2993  
253 QY -----TK-----D----- 255  
2994 Db VFCGVDMNLIANIFTPLVQPVGALDVSASVAGGIIAILVTCAAYYFMKFRFRVFEYNH 3053  
256 QY -----G-----N-----LA-----S--- 260

3054 Db VVAANALLFLMSFTILCLVPAYSFLPGVSVFYLYLTFTYFTNDVSVFLAHLOWFAMESPIV 3113  
261 QY -----T-----L-----G----- 263  
3114 Db PFWITAIYVFCISLKHCHWFFNNYLRKRVNFNGVTFSTFEAACTFLANKEMYKLKLRSE 3173  
264 QY -----Q-----H--K-----G-----P----- 268  
3174 Db TLLPLTOYRYLALYKXKYFSGALDTSYREAAACCHLAKALNDFSNSGADVLVYQPQTS 3233  
269 QY -----F--AL--K-----W----- 274  
3234 Db ITSAVLQSGFRKMAWPSGKVEGCMQVTCGTTTLNGLWLDLDDTVYCPRHVICTAEDMLNPN 3293  
275 QY -----N-----K----- 276  
3294 Db YEDLLIRKNSHSLVQAGNVQLRVIGHISMQNCILLRLKVDTSNPKTKPKYFVRIPQGTFS 3353  
277 QY -----KG--N-----F-I-L--S-----A 284  
3354 Db VLACNGSPSGVYOCAMRPNHTIKGSLNGSCSGVGNIDYDCVSCYMHMELPTGVHA 3413  
285 QY -----VDK-----TTI-----I-----W-----D----- 294  
3414 Db GTDLEGKFGYFPVDRQTAQAAGTDTTITENVLAWLYAAVINGDRWFLNRFTTLNDFNLV 3473  
295 QY -----H-----TG-----E----- 299  
3474 Db AMKYNIEPLTQDHVDILGFLPSAQGTIAVLDMCAALKELLQNGMNGRTILGSTILEDEFTP 3533  
300 QY -----AK-----Q--O-----F-PF-- 306  
3534 Db FVVROCSGVTPOGKFKIVKGTGHMMLITFLTSLILVQSTOWSLFFPVYENAFPLFTL 3593  
307 QY -----H-----S--A-----PA----- 311  
3594 Db GIMAIACAMLLVKHKGHAFCLFLPLSLATVAYFNVMVMPASVMWMRMTWLEADTSLSG 3653  
312 QY --L-D-----V-D--W-----Q----- 317  
3654 Db YRLKDCMYASALVLLILMTARTVYDDAARRVWTLNMVTLVYKYVYGNALDAQISMWAL 3713  
318 QY -----SN-----NT----- 321  
3714 Db VISVTSNYSGVTTIMFLARIVFCVVEYYPPLLFTITGNLTQICIMLYVCFLGYCCCYFGL 3773  
322 QY FASC-----ST-----D----- 328  
3774 Db F--CLLENYFRULTGVVDYLVNSTQEPYMNYSQGLLPKPSIDAFKLNILKILGIGKPKCIK 3831  
329 QY -----M-----CI--H----- 332  
3832 Db VATVQSKNSDVKCTSVLLSVLQQLRVESSSKLWAOQVQLHNDILLAKDTTEAFKMWSL 3891  
333 QY -----V--CK-----L-----G-- 337  
3892 Db LSVLLSMQGVADINRLCEEMLDNRATLOAIASEFSSLPYAAVATAQEAYEQAVANGDSE 3951  
338 QY -----Q----- 338  
3952 Db VVLKLLKSLNVAKSEBFRDAAMQKLEKMAQOMTOMYKQARSEDKRAKRVTSAMQTMFL 4011  
339 QY -----D-----R--P--I-----K-----TF----- 345  
4012 Db TMLRLKNDLANNIINNARDGCVPLNIIFLTTAAKLMVWVVDYGTYKNTCDGNTFTYASA 4071  
346 QY -----Q-----CH-----T-----NE--V----- 352  
4072 Db LWEIQVVDADSKHVQLSEINMDNSPNLAWPLIVTALRANSVAVKLQNNELSPVALROMSC 4131  
353 QY -----NA-----I-----KW-----D-----P-- 359

Db 4132 AAGTTTACTDDNALAYNNNSKGRFVLALLSDHQLKWARPPKSDGTGTIYTELEPPCR 4191  
QY 360 ---T---G-NL---LA---S-CS--- 368  
Db 4192 FVTDTGPKGVKYLFIKGLNNLRGVLGSLAATVRLQAGNATEVPAHSVLVSFCAPAV 4251  
QY 369 D---D---M---T--- 372  
Db 4252 DPAKVDYLASGGQPTINCVMKLTHTGTGQAITVTPKANMQESFGGASCCLYCRCHI 4311  
QY 373 ---LK---I---W---S---M--- 378  
Db 4312 DHPNPKGFCDLGKYQIPTTCANDPVGFTLRNTVCTVCGMMKGYCGSCDQLREPLMQSA 4371  
QY 379 ---K---Q--- 380  
Db 4372 DASTFLNRVCGVSAARLTGCGTGTSTDDVYRAFDIYNEKVAGFAKELKNCRCRFBKDEE 4431  
QY 381 ---D---N---C---VHDL---Q--- 388  
Db 4432 GNLLDSFVVKRHTMSNYQHEETIYNLVKDCPAVAHDPKPRVDGDMVPHISQRRLTKY 4491  
QY 389 ---A-H---N---KEI--- 394  
Db 4492 TMADLVYALRHFDEGNCDTLKEILVTYNCDDDDYFNKKDWDYFVENPDILRVYANLGERV 4551  
QY 395 ---Y--- 395  
Db 4552 RQSLKTVQPCDAMDAGIVGVUTLQNDLGNWYDFGDEQVAPGCGVPIVDSYSLLM 4611  
QY 396 ---T---IKW--- 399  
Db 4612 PILTLRALAAESHMDADLAKPLIKWDLKYDFTEERLCLFDRYFKYQDQYHPNCINCL 4671  
QY 400 ---S---PT---GP--- 404  
Db 4672 DDCILHCANFNLFSTVFPPTSGPLVRKIFVDGVPFVSTGYHRELGVHNDVNHLH 4731  
QY 405 ---G---TNN---P--- 409  
Db 4732 SSRLSPKELLVYAADPAMHAASGNLLDKRTTCSVAALTNNAVAFQTVKPGNKNDFYDF 4791  
QY 410 ---NA---NL-M---L--- 415  
Db 4792 AVSKGFKGSSVELKHFPAQDGNAAISDYDYRYNLPMTCDIRQLLFVVEVDKYFDC 4851  
QY 416 ---A---SASF---DS--- 422  
Db 4852 YDGGCINANQVIVNNLDKSGAGFPFNKGRKARLYDSMSYEDQDALFAYTKRNVIPITILM 4911  
QY 423 ---TV---R---L---W--- 427  
Db 4912 NLKVAISAKNRARTVAGVISCSTMTNRQHQKLLKSIAATRGATVIGTSKFYGGWHNML 4971  
QY 428 ---DV---DR--- 431  
Db 4972 KTVYSDVEPHLMGWDYPCDRAMPNMLRIMASLVLRKHNTCCNLSHRFLRANCAQV 5031  
QY 432 ---G---IC---I--- 435  
Db 5032 LSEMVMCGSLYKPGTSSGDATTAYANSVFNICQAVTANVALLSTDGKNKIADKYVRN 5091  
QY 436 ---H---TLT-KH--- 441  
Db 5092 LQHLRYECLYRNRDVEHFEFYALRKHFSMMLSDDAVWCYNSNYAAQGLVASIKNF 5151  
QY 442 ---Q---E---P---VY---S-V-A--- 449  
Db 5152 KAVLYIQNNVFMSEAKWTETDLTKGHEFCQSQHTMLVKQGDYVLPDPDRRIILGAGC 5211  
QY 450 ---F-S---P---D---G--- 454  
Db 5212 FVDDIVKTDGTLMIERFVSLAIDAYPLTKHPNQEVADVFLYLQYIRKLHDELTHMLDM 5271

QY 455 ---RY---L---A---S---GS---F--- 462  
Db 5272 YSVMLTNDNTSRWBEFYEAMYPHTVLOAVGACVLCNSQTSRLRCGACIRRFELCKCC 5331  
QY 463 -D---K---C---V---H--- 467  
Db 5332 YDHVISTSHKLVLSVNPVVCNAPGCDVTDVTQLYLGMSYCKSHKPPISFPLCANGQVF 5391  
QY 468 ---I---W---NT--- 471  
Db 5392 GLYKNTCVGSNDVDFNALATCDWNTWAGDIILANTCTERLKLFAAETLKATEBTFKLKSYG 5451  
QY 472 ---Q---T---G-A--- 475  
Db 5452 IATVREVLSDRELHLSWEVGKPRPLNRNVVFTGYRVTKNKSKVQIGEYTFEKGDYGDVV 5511  
QY 476 ---L---V---H---S---Y-R--- 481  
Db 5512 YRGTTYKLVNGDYFVLTSHTWPLSAPTLVPOEHYVRITGLYPTLINISDEFSSNVANYQ 5571  
QY 482 ---GTG---G--- 485  
Db 5572 KVGOMKYSTLQGGPGTGKSHFAIGLALYPSARIYVTACSHAADVADALCEKALKYLPIDKC 5631  
QY 486 ---F---E--- 488  
Db 5632 SRIIPARARVECFDKFNKSTLEQYVFCVNALPETTADIVVDFEISMATNYDLSVUNAR 5691  
QY 489 ---VC--- 490  
Db 5692 LRKHHVYIGDPAQLPAPRTLLTKGTLEPEYFNSVCLRMKTIQPDMLGTCRCPPAEIVD 5751  
QY 491 ---W--- 491  
Db 5752 TVSALYVDNKLKAHKDKSAQCFKMFYKGVITHDVSSAINRPOIGVVREFLTRNPAWKAV 5811  
QY 492 ---NA-A---G-D--- 498  
Db 5812 FISPYNQANAVASKILGLFTQTVDSQSGSEYDVIFQTOTETAHSCNVNRFNVAITRAKI 5871  
QY 499 G---AS-A---S---D---G---SV 507  
Db 5872 GILCIMSRLDYKQLQFTSLIPEPRNVATLQANVTGLFKDCSKIITGLHPTQAPTHLSV 5931  
QY 508 ---CV---L-D---LR---K 514  
Db 5932 DIKFKTEGLCVDIPGIPKDMTYRRLISMGMGFX 5963

## RESULT 13

Q6XA09 PRELIMINARY; PRT; 7191 AA.  
ID Q6XA09  
AC Q6XA09;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nonribosomal peptide synthase.  
GN Name=NRPS1;  
OS Alternaria brassicae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;  
OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.  
OX NCBI\_TaxID=29911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14727058;  
RA Guillemette T., Sellam A., Simoneau P.;  
RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria  
brassicae and flanking genomic sequences.";  
RL Curr. Genet. 45:214-224(2004).  
CC -I- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme  
family.  
DR EMBL; AY246697; AAP78735.1; -.

DR InterPro; IPR010071; AA adenyl dom.  
DR InterPro; IPR009081; ACP like.  
DR InterPro; IPR000873; AMP-bind.  
DR InterPro; IPR001242; Condensatn.  
DR InterPro; IPR000977; DNA ligase.  
DR InterPro; IPR006162; Pplantn S.  
DR InterPro; IPR006163; Pp bind.  
DR Pfam; PF00501; AMP-binding; 4.  
DR Pfam; PF00668; Condensation; 9.  
DR Pfam; PF00550; PP-binding; 6.  
DR PRINTS; PR00154; AMPBINDING.  
DR TIGRFAMs; TIGR01733; AA-adenyl-dom; 4.  
DR PROSITE; PS00075; ACP DOMAIN; 7.  
DR PROSITE; PS00455; ACP BINDING; 2.  
DR PROSITE; PS00697; DNA\_LIGASE A1; UNKNOWN 1.  
DR PROSITE; PS00012; PHOSPHOPANTHETINE; UNKNOWN 3.  
SQ SEQUENCE 7191 AA; 793216 MW; F908905155F2394C CRC64;

Query Match 71.4%; Score 2618.9; DB 2; Length 7191;  
Best Local Similarity 8.6%; Pred. No. 3e-41;  
Matches 451; Conservative 48; Mismatches 11; Indels 4751; Gaps 360;

Qy 1 M-----S-----IS-----S-----D-----E----- 7  
Db 1521 MAFRLAGKGRAPAGHAETLRLWASVLKLDAGVGAEDSFRLGSGDSISAMKLVTAAR 1580

Qy 8 ---V-N---F---L---V---Y---R--- 14  
Db 1581 KGVVLNVASVFAQPKLAWATAVMLPSDDAAKPEADTLPMELLPASSRQAVALAASE 1640

Qy 15 -----Y---LQE----- 18  
Db 1641 CDVFPDCIEDMYFCSKQEGVLMNTKDPGYVVQPIYRLPADMDISRFEAMKVVAAE 1700

Qy 19 -----S---GF-----S---H---S-----A-FT----- 27  
Db 1701 ASLRTRIVYSEEGFLQVVVREDIKWQSLPDIQHINETTRRLPAKNGAPLATFTLVGENT 1760

Qy 28 ---F---G---G---S---H---S---P---AA---L--- 49  
Db 1761 DSMFFVWTAHVYDGSWTALFRKVEANYRSGVHGMPTVPYRSFVKYISSLDQRQSDA 1820

Qy 38 -----NI-----NGAL---V---P---P---AA---L--- 49  
Db 1821 FWLTQLDNVTAQFPQPSDPHRVEANGQLLHVLTTRNEGLEVTVPVMIRAAWGILLAT 1880

Qy 50 -----I-SII-----OK----- 55  
Db 1881 YSGSDDDVIMGETNSGREASVPGIESIIGPTITTAAPVRLNRLNRSITVHDYKLTQKQSSLS 1940

Qy 56 -----CLO-----Y-VE---AE---VS---I----- 66  
Db 1941 LPYQFAGLQHIKSLSPETAIACFQSPGLGIEAGDDFPAESSLWNWVSANTIGTDFSYAF 2000

Qy 67 -----N---E-----D---GT----- 71  
Db 2001 VFCKVNAQVQVEALPDDRVTQTLAQRVQVQDFLTQLQNGTDNLRSLDLDLNLPA 2060

Qy 72 -----L-----F-----D----- 74  
Db 2061 DRKTISSWNSKVPPIIPRCIHVSIAEDQTAIRPTACAIDAWDTGFMYSYRLDERSSALAH 2120

Qy 75 -----G-RP-----I----- 78  
Db 2121 QLIRLGVRPKQFVPLCFDKSGWTIVAIIVLKGAFAVPLDFEAPVLRLEIVSDIDADL 2180

Qy 79 -----E---S-----L-SL----- 83  
Db 2181 LLCAPQRELQCSIPCSMTVDRQATETVAGRLPSLVHSDSPAYAFYTSGTGPKGA 2240

Qy 84 -----I-----DA---V---M-----PD----- 90  
Db 2241 VVHTHWTSSTAFAGWKISTASRVLQFASYPDACLIEVFSTLMGGTVCVPCDQGRSRT 2300

Qy 91 -----VW-----O----- 93  
Db 2301 NDLVGVNRPVNWAAALTTSVVRMIVPSEIPQLETFLYGEAMSQDQDLVTWADKYNLNG 2360

Qy 94 -----T-----R----- 95  
Db 2361 YGPTCAAVATSNIMTPHMRPNLGRAVTARGWIVSRNNHTTLPVGAIGELLEGGA 2420

Qy 96 -----O---QA-----Y-R-D--- 101  
Db 2421 AGYLNPNPEKTAQVFVQCARWCVGLMGDDISAPVRIYKTDGLVKYNEDGTMLYLGRKDLQ 2480

Qy 102 -----KL---AQ-----O-----H----- 107  
Db 2481 KVRGQBLESEVEKLLDDHVMQSALASVPTTGPCAKRLVAIVSLQXHGHDHDTDKQLR 2540

Qy 108 -----A---A-----A-----A-----A-----A 113  
Db 2541 LLPQENASLNIAITRDGLCERLPAYMIPSLWIAVERFPLMPSPGKMDRRCAIQMLEQMDQA 2600

Qy 114 -----A---A-----A---A---T---NQ----- 120  
Db 2601 TYRLISAMGTDAAEVNNGSPIEKLRATFAKVLNLNVADVRLNQSFHLGSGDSIAAMQVS 2660

Qy 121 -----QG----- 122  
Db 2661 SQCRAGQFPISVQDIIRSKSISMASAVDLSQSSQVTTAKDYDLPDLSPIQKVPF 2720

Qy 123 -----SA-----KN-G-----E----- 128  
Db 2721 GDKHYFNQTELPRLSRNIETEELRSALTVLAKTHPMLRARFSEAGAWKRIEKOVS 2780

Qy 129 -----NTAN----- 132  
Db 2781 SFLRHHHVQAGNDANLRPIIDHSQATLDVAKGPTFAIELFDVDDTFSSAIALVAHLII 2840

Qy 133 -----G-----E----- 134  
Db 2841 DVVSGWILLEDLQGLQLOPAPQSPMPYHMLHEQSLQATQESARRVFPVGDIA 2900

Qy 135 -----E---NG-A---H-----T 140  
Db 2901 WAMEGRPNVNGDVVEEDHLSTRDTMLLGAQDALATETIDILIAALFESFRKVPDRST 2960

Qy 141 IA-----NN-----H-----TDM----- 148  
Db 2961 IATHNEGHGRETFNNRQDLSRTVGWFSTVTPHLPVPLNEATDMISTIRWVDRFRNRTPD 3020

Qy 149 -----ME---V-----DG-----DV-----E-I 157  
Db 3021 KGRPYFAYRNLTEEGOTRPFASHWPAEAVPNVGRLOSQDNKQGLFTALSVDVDSREV 3080

Qy 158 P-----S-----N-----K-----AV-----V 164  
Db 3081 PRLALFDITAAVSQGAIKLSFGWNRNRQKEIRAWVGKCRQTLVDAVEELLQARQERSV 3140

Qy 165 -----L---R-----G-----H-E----- 169  
Db 3141 GNFKYLLLYNGISRLSAILPAGINLNDVEDIYPASPMOQGLLHTQSRHPELYTYHTVSQ 3200

Qy 170 --S-----E-----VFI----- 174  
Db 3201 VQSADGNPIDPRLAQWQVVVHRHQALRTIFIDSLAKGSKDQIVLKEKGRVQIILADC 3260

Qy 175 -----C-A-----W-----N----- 178  
Db 3261 DDSQVANLLRHQSSIDCREALPBRMSICKTKGRVWFKLESHVINDGTSVSNLLADLA 3320

Qy 179 -----P---V-----SD-----L---V---SG-----S 188  
Db 3321 RAYARKLTRADAGPLYSDYIGYMLSRSSDADLAWKAHLSGIEFCLFPVLNDGIPSPES 3380



189 G-D---STARI-----W-----N 197  
3381 GSDVELGSTSRVQDFCKONGVTLNVLQTLWALTJLHYVGTDFDVSFGLIASGRDIPVTN 3440  
198 LSE-----N-----ST---S---G-----P-----206  
3441 INEAVGCFVMVVSRLSFSDETTIAQLLEALQGTSEALSHOGCSLADIQHALQLPSLFN 3500  
207 ---T-Q-----LVL-----R-----212  
3501 TATFQERSLSSDPEDTALIYEDMEAEADAGEYIVTVNADVTQSIIVDFGYSKDRILPSQ 3560  
213 -----H-----C-----I--215  
3561 AQNAETFFKILDSIVVCSASELTIGKLDILTTESSIHQIMENPQLPPPIRRCLHDVHD 3620  
216 ---R---EG---G---QD-----221  
3621 QALTRPRTTKAVEGWDGTFYQDFDKITNQLAVHLOSIGVTTTETFPILPEKSSYAIVSM 3680  
222 -----221  
3681 IAIMKAGGAYVLPDPKHPQTRRELIEDVGSVVLCSRGVHTTASEVAKTAVIVDORSIR 3740  
222 ---VP---S-----N-----K-----226  
3741 KLGVPISSKPRTCATPDNAAAYCLFTSGTCKPKGTIIPHQAFCFSAAAFTRMNINATSR 3800  
227 ---D---VTS-L-----D-----W-----233  
3801 TFQFASYTFDASCIEILSALTVGATVCVPTEDDRMNAAGAIRKLRVNMSSLTPSVLGTI 3860  
234 -----N-----S 235  
3861 EPERVPLKTLVSGGEALSGPIKKWSNSTCFINAYGPTCSVVAATAYKSTLDHKLIVS 3920  
236 E-GTLATGS---Y-----D-----G---FAR-----I---251  
3921 EPGT-IGTSGCLWIVHPRNHDKLMPVSGVELVIEGPTVARGYLNDVKTAKAFINDP 3979  
252 -W-----T---K-----DG--NL---AST---L--GO-----H--265  
3980 AWAKTIFSSNNTFEAARMYKTGDLVRYNTDGSVNYIGRKDTQIKLANGQRIELGEIBFHV 4039  
266 -----265  
4040 KNPPERVQSAVELVAPSNRSSAKALAVFAVVOQDAIDGQSVQVQASTDLPAADLLLLPL 4099  
266 ---K-----G-----P-IF---ALKW-----274  
4100 SDELRDMCKNTENGLAGSLPSYMPIAIFIPVTKLPWTSAGKLDNRRLSLVQNLBRETA 4159  
275 ---NKK-----G-----278  
4160 MYRLSTANKKPEITEAEKKIHKAVCSVLSPSSVGIDDSFVRLGGDSISSMRLVAMAH 4219  
279 -----N-----P-----I-----281  
4220 TEQMBLSFDIFKNPKLSDIAKTAQISKSQAQEKVQMPDPLLPASLTRSDVISEVVQOC 4279  
282 ---LS---AG---V-----DKT-----289  
4280 QVSKEDLQDAYPTSSLODALLTSLIKQAGAYVAQHVLAFLPKSLDMTKFKAAWESATQEID 4339  
290 ---T-II-----W-----DA-----H-----T---297  
4340 ILTRITQPSGIFMOAVLRNDPVDWREAKSLKSAEDDASKIPPHLGHLAAVTLVTTS 4399  
298 GE-----A-----K-----Q-----302  
4400 GERYFVWTLHLYDGSIVYLMQVQOQIYKSGVSTTPQTSYARFVEYLSSTSVSDSVY 4459  
303 ---QFP-----P-HS---A-----PA-----311

4460 WRERLTGVNAVQPRPRSHATSSAPPNGQMFQHSKMIHAKVNTDVTANARAAWALILAA 4519  
312 ---L-----DV-----314  
4520 YTGSDDDVFGETLAGRDVAMTGITDVGPTLTTPVSRVKIDRGATVSDLLNTIATNITDR 4579  
315 -----D-W-----Q-----317  
4580 IPYQHGLSNAIKALGEDMIAACDFQNLVLVQIOTENEELADSMWSVHDNEEOGNFTYPLVI 4639  
318 ---S-----N-----319  
4640 ECKMGLSKTEVLAHPDANVISLHWVORLIYQFETVLIQSATHVRHIAVLSQDQKLVR 4699  
320 --N-----T---F-----AS-----324  
4700 KMNAYEPRLIDTVPSPFPKKAASQPTTAVTAFDGEFSYGELSALASQAQLVLPFGAG 4759  
325 ---C-----S-TD-----328  
4760 PECLIPICVDKSRWAIVAILAILISGAGYVPLSPDTPASRHLHIVETCKASIVLCSPKYT 4819  
329 ---M-----C-----I-H---332  
4820 HRFVEMVGHVFSVSETAIRQLPTSSISLSORAKGNICVIFTSGTGLPKGVVIEHKS 4879  
333 ---VCKLG-----Q-----D-----R-----P-----341  
4880 VSSAAACE-GLHITPSTRVFOFCFLFDVSVGETLTVLLRGATICVPSDEQRTNLAAA 4938  
342 IKT---F-----QG-----HTN---350  
4939 V-TDLNANWAFLTSPVASTLEGKPSVPTLETLLVVGGEAMTSDVVVKWATGVNLH-NGYGP 4996  
351 -E-V---N---A---I---K---W---D---P---T---360  
4997 TEGTVAIGNDHVSAQRDPSNIGHPLKSGRAWLTNSDNPHELAPIGATAECLLEGPLLAR 5056  
361 G-N-----L-----L---A---S-----366  
5057 GYLNDPKRTSEAFIEAPAFKPNFSSNESRIYRTGDLVQVADGSIQYMGKRKNQIKLAG 5116  
367 -----C-----SD-369  
5117 QRIELDEIQAVHADNNHVQVVQVLPKVGPCYKTLTVVVSFPCTAASAGSDRRILSDT 5176  
370 -----D---M-----TL-----K-----374  
5177 ESLSQINRRARDRLADLVPSYVVPFIWIAVPRIPTLASAKLDKKQVGLWLEGMDALYQRI 5236  
375 -----IW-----S-MK-----379  
5237 MGAELPEDMEGFAAALTVLRGIAWAKVLRNPVEDVKPSKSWLSLGGDSISAMKLLAKCRS 5296  
380 -----Q-D-----381  
5297 EGINLNLQILRAKSLSHLAADVKSVVILDHEQNDRFALSPIORFYVEAGSIENSTH 5356  
382 -----N---C-----V-----384  
5357 FNOGSTLRILRVQPAWQVQALNSIVECHSMLRAREFSKDNNGQWQVMSKVSQSYAFTA 5416  
385 HD-----L---Q---AH-----390  
5417 HDVSTASAAGAIISSTQKSLDIRTGPFAVDLFNLKQHQLFMAAHLHLVDDVVSWGILLG 5476  
391 -----390  
5477 DLEDLGSQVPTTLPRSLPFTQWCKMQTSASBITOQLTVKNOPLVVEPAPFAMGDVDR 5536  
391 -N-----KE-----IY-T-----I-K-----398

Db 5537 PNVYCDSEDFVIDKETSAMAFDNHHVYKTDLDVILLAAILHSESRVFINKRAPTLENE 5596  
QY 399 -----W-----S-----P-----T-----G-P-- 404  
Db 5597 SHGREVDGNSLDLSRTVGWFTLYPVTPIDDEDEVIHTRQVKDTRRKVASNGRPYF 5656  
QY 405 -----G-----TN-----N-----P----- 409  
Db 5657 AHRVLTEDGKERFANHAPMEVLFNYLGRQESGSHDSLSTQVEGDDDETSVGVKTSR 5716  
QY 410 -----N-----AN-----L-----M- 414  
Db 5717 MALFEISASVTEGOIQLSFMWNRYSKKQKIRRWIABECORTLEIAIELAKTRDPQTM 5776  
QY 415 -----LAS-----A-----SPD-----ST- 423  
Db 5777 DFPLPLESRLDRVLKTLPHAGVPSFQVEDMVPSPIDGMILSQIKSPESYWSSTT 5836  
QY 424 --VR-----L-W-----D-----V-DR-----GI-----C 434  
Db 5837 FEVRSKRGPVDASKVVDGKQVVARHPALRTIFIDSVCKGVPDQIVVKSPDSGIVTYKC 5896  
QY 435 -----I-HT-----L-----T-----K-----H----- 441  
Db 5897 ADAELATLLESIRHSSSLGKKPVLPQAAVVOQTSSGKIFVKVIVNHAVIDGSGSLVIGQ 5956  
QY 442 --QE-----PVVS-----VA-----F-----S-P-- 452  
Db 5957 DLQAYEGRUSEDGLYSAYIKYLRALPAEDAIAYKAKLRGVSPCYFPTTPRDPSPRQ 6016  
QY 453 -----D-----G-----R-----YLAS----- 459  
Db 6017 LOSLDMRFTFDELHDLAESNVNTIANILAAWALVLRSTNSDVCYGLTSGRNVPID 6076  
QY 460 -----GSF-----D-----K-----C----- 465  
Db 6077 RIENAVGAFINMLVSRTELKATSLSLEIVENVQSDFVGSMPHQCSLAQFOHDLGLSGKS 6136  
QY 466 -----V-----H----- 467  
Db 6137 LFNTAVSIQRSSPELTSDSGIEFQDLGHPSEFAITWNIDATRDGEVFTYWSDA 6196  
QY 468 ----- 467  
Db 6197 VTDEAKNVSTLMAKILVQLSNPKQTIAELDIVVKGPAQPAVSRNLNPKPRPSILRSS 6256  
QY 468 -----I-----W-----N----- 470  
Db 6257 SSISRSTSPRTPRTFPDLPAPALPAETPDWSSLIRSIWSEWVPQIVEQIVAKNLS 6316  
QY 471 -----TQ-TGAL-----VHSYR----- 481  
Db 6317 TEPTSATIDQMTQMTGNLTKRTSISORGRPSIDDDAGSPRAASIHSSRAMSIASNAENRI 6376  
QY 482 -----G-----T-----GG--I 486  
Db 6377 QTAADMVATLGLVLAETASKVADPVEKLLNLAELLENVEETIEQDDSFQGLGDSII 6436  
QY 487 -----F-----EV-C----- 490  
Db 6437 AMRLVGAAREGLSMTVADVFNKPTFADMARVVRVAGEVIDEVMSRAGGSVAGRSAGQS 6496  
QY 491 -----W-----NA----- 493  
Db 6497 RSLRHDRAPIWSEFQDIMSDDPNADTKSIAPSEMPAEPGLNSRDSMTFQWQGLTNNQAR 6556  
QY 494 -----A-----GD-----KV-----G----- 499  
Db 6557 PAASREVSKSLATHITIQEGVESSVNRVSMLGDPNVESVISKVQVKGKGISDVFPVTFDF 6616  
QY 500 --A-----S-----A----- 502  
Db 6617 QSLAITGTLMESKMWLNLYFYLDGGLDLRLKQLQAAYRMVHAFDILRTVFVPYGRFLQV 6676

QY 503 -----SD-----G----- 505  
Db 6677 VLRLQPEFYHQTDDDDITFTKDLRKQKRENGPKLGEAFIQFVAKQKQOTGRYRIFMRL 6736  
QY 506 S-----VC-----VL-----D-----L-----R---K 514  
Db 6737 SHAQYDGVCMKILNALQDGYNGLPVSSAPSGCNFVRETAK 6777

## RESULT 14

AAP78735  
ID AAP78735 PRELIMINARY; PRT; 7191 AA.  
AC AAP78735;  
DT 01-APR-2004 (T-EMBLrel. 27, Created)  
DT 01-APR-2004 (T-EMBLrel. 27, Last sequence update)  
DT 01-APR-2004 (T-EMBLrel. 27, Last annotation update)  
DE Nonribosomal peptide synthase.  
GN NRPS1.  
OS Alternaria brassicae.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;  
OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.  
OX NCBI\_TaxID=29911;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14727058;  
RA Guillemette T., Sella A., Simoneau P.;  
RT "Analysis of a nonribosomal peptide synthetase gene from Alternaria  
RT brassicae and flanking genomic sequences.";  
RL Curr. Genet. 45:214-224(2004).  
DR EMBL; AY246697; AAP78735.1; -;  
SQ SEQUENCE 7191 AA; 793216 MW; F908905155F2394C CRC64;

Query Match 71.4%; Score 2618.9; DB 2; Length 7191;  
Best Local Similarity 8.6%; Pred. No. 3e-41;  
Matches 451; Conservative 48; Mismatches 11; Indels 4751; Gaps 360;

QY 1 M-----S-----IS-----S-----D-----E----- 7  
Db 1521 MAPFLAGKSGRAPAGHAETVLARLWASVLKLDAGAVGEDSPFRFGDGSISAMKLVTAAR 1580  
QY 8 ---V-N---F---L---V---Y---R--- 14  
Db 1581 KQGVNLVASFAQPKLAMAATAVMLPSDDAAKPEADTLPMELLPASSRQAIVALAAE 1640  
QY 15 -----Y-----LQE----- 18  
Db 1641 CDVFPDCIEDMYPCKLOEGLVMTNKPQTVVQPIYRLPADMDISRPFKEMKVVAAE 1700  
QY 19 -----S---GF-----S---H---S---A-PT----- 27  
Db 1701 ASLRTRIVYSEHGFQVVRREDIKWQSLPDQIHNETTRRLPAKNGAPLATFLVGENT 1760  
QY 28 ---F-----G-----IE---S-H-----IS---QS-- 37  
Db 1761 DSMFPWTAHAHYDGSWTALFRKVEANVRSGVHSPATVPVSRPVKYISSLDQSDA 1820  
QY 38 ---NI-----NGAL---V---P---P---AA---L--- 49  
Db 1821 FMTQLDNTAAQFPOLPSPDHRVEANGQLLTLTRNPGLEVTVPVSMIRAAWGILLAT 1880  
QY 50 -----I-SII-----OK----- 55  
Db 1881 YGSSDDVINGETNSGREASVPGIESIIGPTITAPVRLNRLSLTVHDYKQKQSSLS 1940  
QY 56 -----GLQ-----Y-VE-----AE-----VS---I--- 66  
Db 1941 LPYQFAGLQHISKLSPEAIACDFQSFGLGIEAGDDFDASSLWNVMSANTIGTDFPSYAF 2000  
QY 67 ---N---E-----D-----GT----- 71  
Db 2001 VFNCNKVATGVQVEALFDDRVRVQVTLAQRMVQOQDFILTQLNGTNIILRSLODLDLINPA 2060

QY 72 -----L-----F-----D----- 74  
Db 2061 DRKTISSMNSKVPPIIPRCHSVIABDQALRPTACAIADWDGTFMSYRELDESSALAH 2120  
QY 75 -----G-PP-----I----- 78  
Db 2121 QILRLGVRPKQFVPLCFDKSGMTIVAIIVLKAGAFVPLDFEAPVLRIRREIVSDIDADL 2180  
QY 79 -----E-----S-----L-SL----- 83  
Db 2181 LLCAPQYRELQSIQPCSTWVDRQATETVAGRLPSLPSVHSDSPAYFTSGTGPKGA 2240  
QY 84 -----I-----DA-----V-----M-----PD----- 90  
Db 2241 VVHHTHWTSSTAFAGWKISTASRVLPASVTFDACLIEVFSTLMQGGVTCVPOQGSRT 2300  
QY 91 -----VV-----Q----- 93  
Db 2301 NDLVGVINRPNVNAALTPSVVRMIVPSEIPQLETFLVGEAMSOQDLVTWADKVNLSNG 2360  
QY 94 -----T-----R----- 95  
Db 2361 YGTECAAVATSNIMTPHMRPNHNLGRAVTARGHIVSRNHHHTLAPVGAIGELLLEGAVG 2420  
QY 96 -----Q-----QA-----Y-R-D----- 101  
Db 2421 AGYLNNPETAQVVFQARWCVGLMGDDISAPVRIYKTDGLVKYNEDGTMLYLGRKDLQT 2480  
QY 102 -----KL-----AQ-----Q-----H----- 107  
Db 2481 KVRGORLELSEVEHKLDDHVMQSALASVPTTGPCAKRLVAIVSLQKHGDHDTDTDKQLR 2540  
QY 108 -----A-----A-----A-----A----- 113  
Db 2541 LLPQENASLIATIRGCLERLPAYMIPSLWIAVERPPLMPSGKMDRRCAIOWLEQMDQA 2600  
QY 114 -----A-----A-----T-----NQ----- 120  
Db 2601 TYRLISAMGTDDAEVNVNGSPIEKKLRAIFAKVLNLSVADVRLNQSFHLGGDSIAAMQVS 2660  
QY 121 -----QG----- 122  
Db 2661 SOCRAQFPISVDIIIRKSISAMASAVDLSSQSVTTEAKDYDLPDLSPQIKQVFFDAV 2720  
QY 123 -----SA-----KN-G-----E----- 128  
Db 2721 GDKHNYFNQTELPRLSRNIETEELRSALTVLAKTHPMLRARFSGNEAGAWKQRIEKDVS 2780  
QY 129 -----NTAN----- 132  
Db 2781 SPFLRHHVQAGNDANLRPIIDHSQATLDVAKGPTFAIELFDVDDTFSQAIALVAHLLII 2840  
QY 133 -----G-----E----- 134  
Db 2841 DVVSGILLLEDLQGLQLOPQSPMPYHMLHEQSLQTESARRVFPVGDIAPEGDLDY 2900  
QY 135 -----E-----NG-A-----H-----T 140  
Db 2901 WAMEGRPNVNGVVEEDLHLSTRDTMLLGAQDALATETIDILIAALFSSFRKVPFDRST 2960  
QY 141 IA-----NN-----H-----TDM----- 148  
Db 2961 IAIHNEGHGRETTNNRQDLRSRTVGFSTVTPHLPVPLNEATDMISTIRWDRFRNRTPD 3020  
QY 149 -----ME-----V-----DG-----DV-----E-I 157  
Db 3021 KGRPYFAYRNLTTEGQTRFASHWPAEAVFNVYGRLOSQDNKQGLFTALSDVDSREVGEDV 3080  
QY 158 P-----S-----N-----K-----AV-----V 164  
Db 3081 PRALFDITAAVSOQAIKLSFGNRMNRQKEIRAWVGKRCQTLVDVAVEELLQARQERSV 3140  
QY 165 -----L-----R-----G-----H-E----- 169

Db 3141 GNFKYLPLLYNGISRLSAILPAGINLNDVEDIYPASPMOQGLLHTQSRPELYTYHTVSQ 3200  
QY 170 -----S-----E-----VFI----- 174  
Db 3201 VOSADGNPIDPRRLAEAWQVVRHQALRTIIFIDSLAKDGSKDQIIVLKEKPRVQIADLC 3260  
QY 175 -----C-A-----W-----N----- 178  
Db 3261 DDSQVANLLRHQSSIDCREALPPHRMSICKTKTGRVWFVKLELSHVINDGTSVSNLLADLA 3320  
QY 179 -----P-----V-----SD-----L-----L-----V-SG-----S 188  
Db 3321 RAYARKLTRADAGPLYSDYIGYMLSRSSDADLAYKHAHLSGIEPCLPVPLNDGIPSPES 3380  
QY 189 G-D-----STARI-----W-----N----- 197  
Db 3381 GSVDELGSTRVQDFCKONGVTLSNVLTWALTUHYVGTGDFVSFGLIASGRDIPVTN 3440  
QY 198 LSE-----N-----ST-----S-G-----P----- 206  
Db 3441 INEAVGCFVNVVSRLSFSDETTIAQLLEALQGTSTEALSHQGCSLADIQHALQLSFLN 3500  
QY 207 -----T-Q-----LVL-----R----- 212  
Db 3501 TAFTRQSRSLSDPEDTALIYEDMEADAGEYIVTVNADVTQDSITVDFGSKDRILPSQ 3560  
QY 213 -----H-----C-----I----- 215  
Db 3561 AQNMATFKIILDSIVVCSASELTIGKLDILTESSIHQIMWNPQLPPPIRRCLHDVHID 3620  
QY 216 -----R-----EG-G-----QD----- 221  
Db 3621 QALTRPRITTKAVEGWDGTFTYQDFDKITNQLAVHLQSIGVTTETFPVILPEKSSYALVSM 3680  
QY 222 -----N-----K----- 221  
Db 3681 IAIMKAGGAYVPLDPKHPQTRLRELIEDVGASVVLCSRGVHTTIASEVAKTAVIVDQSRIR 3740  
QY 222 -----VP-S-----N----- 226  
Db 3741 KLGVPISSKPRTCAFDNAAAYCLFTSGTGKPKGTIIIPHOAFCTSAATFRRMNINATSR 3800  
QY 227 -----D-----VTS-L-----D-----W----- 233  
Db 3801 TFQFASVTTDASCIEILSALTVCATVCVPTEDDRMNAAGAIRKLVNMSLLTSPVLGTI 3860  
QY 234 -----N----- 235  
Db 3861 EPERVPLKTLVSGGEALSGPILKKWSNSTCFINAYGPTCSVVAATAVYKSTLDHKLIVS 3920  
QY 236 E-GTLLATGS-----Y-----D-----G-FAR-----I----- 251  
Db 3921 EPGT-IGTSGCRLWIVHPNRHDKLMPVSGVSELVIEGPTVARGYLNDEVKTAFAINDP 3979  
QY 252 -W-----T-----K-----DG-NL-----AST-----L-GQ-----H----- 265  
Db 3980 ANAKTIFSSNNTFEAARMYKTGDLVRYNTDGSVNYIGRKDTQIKLNGQRIELGEIEFHVG 4039  
QY 266 ----- 265  
Db 4040 KNPPERVQSAVELVAPNSRSSAKALAVFAVVDQDAIDGEOVSQVQASTDLPAAADLLPL 4099  
QY 266 -----K-----G-----P-IF-----ALKW----- 274  
Db 4100 SDELDMCKNTENGLAGSLPSYMPIAFTVTKLPWTSAGKLDNRNLSRVQNLSETMA 4159  
QY 275 -----NKK-----G----- 278  
Db 4160 MYRLTSIANKKPKITEAEKKHAKVCSVLSPSSVIGDIDSFVRLGGDSISSMRLVAMAH 4219  
QY 279 -----N-----F-----I----- 281

Db 4220 TEOMELSFIDIFKNPKLSLAKIQAQISKSQAEKVMQPFLLPASLTRSDVISEVVQOC 4279  
QY 282 ---LS---AG---V---DKT--- 289  
Db 4280 QVSKEDQDAYPTSSLODALLTUSIKAGAGVAVQHVLAALPKSLDMTKFAKAWESAIOEID 4339  
QY 290 ---T-II---W---DA---H--- 297  
Db 4340 ILTRIIQMPGIFMQAVLRNDPVDWEAKSLKSAEDDASKIPPHLGCHLAAVTLVTTPS 4399  
QY 298 GE---A---K---Q--- 302  
Db 4400 GERYFVMTLHLYDGSIVLMQVQOYVYKSGVSTTPQTSYARFVBYLSSTSVSDSVY 4459  
QY 303 ---QRP---F-HS---A---PA--- 311  
Db 4460 WRRLTGNAVYQPPRPSHATSSAPPNGQMFQHSKMAHAKNTDVTAPANARAUALILAA 4519  
QY 312 ---L---DV--- 314  
Db 4520 YTGSDVVFGETLAGRDVAMGTIDVCGPTLTVVPSVKIDRGATVSDLLNTIATNITDR 4579  
QY 315 ---D-W---Q--- 317  
Db 4580 IPYQHGLSAIKALGEDMIAACDFONLLVIQTENEELADSMWSVHDNEEQGNFTYPLVI 4639  
QY 318 ---S---N--- 319  
Db 4640 ECKMGLSKTEVLAHFANDVLSLHVQRLLIYQFETVLQLOSATHVRHIAVLSDQKQLVR 4699  
QY 320 ---N---T---F---AS--- 324  
Db 4700 KKNAYEPRLIDDVPSLFFPKKAASQPTTAVTAFDGEFSYGELSALASQLAQELVKFAG 4759  
QY 325 ---C---S-TD--- 328  
Db 4760 PEGLPICVCKRWAI VAILAILISGAGYVPLSPTDPASRHLHIVETCKASIVLCSPKYT 4819  
QY 329 ---M---C--- 332  
Db 4820 HRFVEMGVHVSSETAIRQLPTSSLSORAKNNICYIFTSGTGPKGVVIEHKS 4879  
QY 333 ---VCKLG---Q---D---R---P--- 341  
Db 4880 VSSSAALCE-GLHITPSTRVFCFSFLFDVSVGETLVLRGATICVPSDEQRTNLAA 4938  
QY 342 IKT---F---QG---HTN--- 350  
Db 4939 V-TDLNANWAFLPVASTLEGPKSVPTLETLVVGGEAMTSDVVDKWATGVNLH-NGYGP 4996  
QY 351 -E-V---N---A---I---K---W---D---P---T--- 360  
Db 4997 TEGTVFAIGNDHVSAQRDPNSIGHPLKSGRAWLTNSDNPHELAPIGATBELCLEGP LLAR 5056  
QY 361 G-N---L---L---A-S--- 366  
Db 5057 GYLNDPKRTSEAFIAPFLNFSSNESRIYRTGDLVQYAADGSIQYMGKKNQIKLAG 5116  
QY 367 ---C---SD--- 369  
Db 5117 QRIELBEIQAVHADNNHVQVQLPKVGPCTKKLTWVVSFGPTAASNAGSDWRRILSDT 5176  
QY 370 ---D---M---TL---K--- 374  
Db 5177 ESLSQINRARDRLADLPVSYNVPFIWIAVPRIPTLASAKLDKQVGLWLEGMDREALYQRI 5236  
QY 375 ---IW---S-MK--- 379  
Db 5237 MGAELPEDEMGPGNAALTVLRGIVAKVLPVPEDEVKPSKWSLGGDSISAMKLLAKCRS 5296  
QY 380 ---Q-D--- 381  
Db 5297 EGINLNLQILRAKSLSHLAADVKSVVILDHEQNDRPPALSPIORFYVEAGSIENSTH 5356

QY 382 ---N---C---V--- 384  
Db 5357 FNOSTILRLRYVQPAWQALNSIVECHSMLRPARFSKNNQWQOQLVMSKVSYSYAF TA 5416  
QY 385 HD---L---Q---AH--- 390  
Db 5417 HDVSTASAAGAILSSQKSILDIRTPGFAVDLFLNKGHQLMFLMAAHLHVIDVDSWGILLG 5476  
QY 391 --- 390  
Db 5477 DLEDLGSGPVTLPRLSPFQWCKMQTSASBITQOLITVKNQPLVVEPANFANFAGWMDVR 5536  
QY 391 -N---KE---IY-T---I-K--- 398  
Db 5537 PNVGSEDERDFVIDKETSAMAFDNHHVYKTLVDILLAILHSFRVFINRKAPTUFNE 5596  
QY 399 ---W---S---P---T---G-P--- 404  
Db 5597 SHGREVWDGNSLDLSRTVGWFTTLYPVTVPIDDEDEVIHTLFQVKDTRRKVASNGRPYF 5656  
QY 405 ---G---TN---N---P--- 409  
Db 5657 AHRYLTEDGKERFANHAPMEVLFNGLROEQSGHSDLSLPTQVEGDDDETSDVGKTSR 5716  
QY 410 ---N---AN---L---M- 414  
Db 5717 MALFEISASVTEGQIOLSFMYNRYSKNQKIRRIWAEACORTLEBIEAIELAKTORPQPTMA 5776  
QY 415 ---LAS---A---SFD---ST- 423  
Db 5777 DPPLLPLESVRLDRVLKTLPHAGVPSFQVEDMYFCSPIQDGMILSQAISPSYSWSTT 5836  
QY 424 ---VR---L-W---D---V-DR---GI---C 434  
Db 5837 FEVRSKRGPDASKVVDGKQVVARHPALRTFIDSVCKRGVFDQIVWKSPPDSGIVTYKC 5896  
QY 435 ---I-HT---L---T---K---H--- 441  
Db 5897 ADDELATILLESINHSNGKKPVLPHQAAVQVOTSSGKIFVKVIVNHAVIDGSSLGWIGQ 5956  
QY 442 ---QE---PVYS---VA---F---S-P--- 452  
Db 5957 DLQEAVEGRLEDGPLYSAVIKYLRALPAEDAIYWKAKLGVSPCVFPTTPRDPSPKPRQ 6016  
QY 453 ---D---P---G-R---YLAS--- 459  
Db 6017 LQSLDMRFRFDELHDLAESSNVVTIANILLAAWALVRSYTNSSDVCYGYLTSGRNVPID 6076  
QY 460 ---GSF---D---K---C--- 465  
Db 6077 RIENAVGAFINMLVSRIELKATSLSLEIVENVQSDPFGSMPHQCHSLAQFOHDLGLSGKS 6136  
QY 466 ---V---H--- 467  
Db 6137 LFWTAVSIQRRSSPEELTSDDSGIEFEQDGDHPSEFAITVNIDATRDDEGVFTYWSDA 6196  
QY 468 --- 467  
Db 6197 VTDEAKNVSTLMAKILVQLSNPKOTIAELDIWVKRPAQPAVSRNLNPKRPSILRSS 6256  
QY 468 ---I---W---N--- 470  
Db 6257 SSISRSSTSPPTPTFPDLAPAAPLPAETPDWSSLIRSIVSEMVQPIVEQIVQVONKLS 6316  
QY 471 ---TQ-TGAL---VHSYR--- 481  
Db 6317 TEPTSATIDMTQTMGLTRKTSISQGRGPSIDTAGSPRAASIHSSRRAMSIASNAENRI 6376  
QY 482 ---G---T---GG---I 486  
Db 6377 QTAADMVATLGLVATEASSKVAPDFVEKKLLNLWAELEWVEETIBQDDSPFQDGSII 6436

```
QY 487 -----F-----EV-C----- 490
Db 6437 AMRLVGAAREGLSMTVADVFNKPTPADMARVVRVGEVIDEVMRSAGGSVAGSAGQS 6496
QY 491 -----W-----NA----- 493
Db 6497 RSLRHDRAPISEFQDIPQDIPNADTKSIAPSEMPAEPGLNSRSDTMFORWOGLTTNOAR 6556
QY 494 -----A-----GD-----KV-----G----- 499
Db 6557 PAASREVQSLSLATHIQEGVESVNRVSMLGDPNVESVISKVQFKGSDVFPVTFD 6616
QY 500 -----S-----A----- 502
Db 6617 QSLAITGTLMSKWLNLVFLYDGDGDLRLKLRQAAYRMVHAFDILRTVPVPGDRFLOV 6676
QY 503 -----SD-----G----- 505
Db 6677 VLKLOPEFIYHQTDDIDFTTKDLRQKORENGPKLGEAFIQFVVAQKQGRYIFMRL 6736
QY 506 S-----VC-----D-----L-----R---K 514
Db 6737 SHAQYDGVCMKSLNALQDNGYGLPVSSAPSGFNVRETAK 6777

RESULT 15
LR3B MOUSE
ID_LR3B_MOUSE STANDARD; PRT; 4599 AA.
AC Q9J1I8; Q8B2D3; Q8B2M7;
DT 05-JUL-2004 (Rel. 44, Created)
DT 05-JUL-2004 (Rel. 44, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Low-density lipoprotein receptor-related protein 1B precursor (Low-
DE density lipoprotein receptor-related protein-deleted in tumor) (LRP-
DE DIT).
GN Names=Lrp1b; Synonyms=lrp1d;
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
RA Yaklichkin S., Lisitsyn N.;
RA "The structure of the mouse homologue of the human candidate tumor
RA suppressor gene Lrp1b/LRP-DIT";
RT Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE OF 3770-4599 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Cerebellum;
RX MEDLINE=22354683; PubMed=12456851; DOI=10.1038/nature01266;
RA Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
RA Nikaide I., Osato N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
RA Baldarelli R., Hill D.P., Bult C., Hume D.A., Quackenbush J.,
RA Schriml L.M., Kanpin A., Matsuda H., Batalov S., Beisel K.W.,
RA Blake J.A., Bratt D., Brusic V., Chothia C., Corbani L.E., Cousins S.,
RA Dalla E., Dragani T.A., Fletcher C.P., Forrest A., Frazer K.S.,
RA Gaasterland T., Gariboldi M., Gissi C., Godzik A., Gough J.,
RA Grimmond S., Gustincich S., Hirokawa N., Jackson I.J., Jarvis E.D.,
RA Kanai A., Kawai H., Kawasawa Y., Kedzierski R.M., King P.B.,
RA Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons P.A.,
RA Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki H.,
RA Nagashima T., Numata K., Okido T., Pavan W.J., Perlea G., Pesole G.,
RA Petrovsky N., Pillai R., Pontius J.U., Qi D., Ramachandran S.,
RA Ravasi T., Reed J.C., Reed D.J., Reid J., Ring B.Z., Ringwald M.,
RA Sandelin A., Schneider C., Sempile C.A., Setou M., Shimada K.,
RA Sultana R., Takenaka Y., Taylor M.S., Teasdale R.D., Tomita K.,
RA Verardo R., Wagner L., Wahlestedt C., Wang Y., Watanabe Y., Wells C.,
RA Wilming L.G., Wymshaw-Boris A., Yanagisawa M., Yang I., Yang L.,
RA Yuan Z., Zavolan M., Zhu Y., Zimmer A., Carninci P., Hayatsu N.,
RA Hirozane-Kishikawa T., Konno H., Nakamura M., Sakazume N., Sato K.,
RA Shiraki T., Waki K., Kawai J., Aizawa K., Arakawa T., Fukuda S.,
RA Hara A., Hashizume W., Imotani K., Ishii Y., Itoh M., Kagawa I.,
```

```
RA Miyazaki A., Sakai K., Sasaki D., Shibata K., Shinagawa A.,
RA Yasunishi A., Yoshino M., Waterston R., Lander E.S., Rogers J.,
RA Birney E., Havashizaki Y.;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs";
RL Nature 420:563-573 (2002).
CC -!- FUNCTION: Potential cell surface proteins that bind and
CC internalize ligands in the process of receptor-mediated
CC endocytosis.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- SIMILARITY: Belongs to the LDLR family.
CC -!- SIMILARITY: Contains 17 EGF-like domains.
CC -!- SIMILARITY: Contains 32 LDL-receptor class A domains.
CC -!- SIMILARITY: Contains 25 LDL-receptor class B domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-aib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; AF270884; AAF81758.1; -.
DR EMBL; AK035795; BAC29188.1; -.
DR EMBL; AK034122; BAC28594.1; -.
DR HSSP; Q07954; 1CR8.
DR MGD; MGI:2151136; Lrp1b.
DR InterPro; IPR011044; Amine_DH_B_like.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_like.
DR InterPro; IPR009030; Grow_fac_recept.
DR InterPro; IPR002172; LDL_receptor_A.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR Pfam; PF00008; EGF; 15.
DR Pfam; PF00057; Ldl_recept_a; 32.
DR Pfam; PF00058; Ldl_recept_b; 33.
DR PRINTS; PR00261; LDLRECEPTOR.
DR SMART; SM00179; EGF_CA; 4.
DR SMART; SM00192; LDLA; 32.
DR SMART; SM00135; LY; 36.
DR PROSITE; PS00010; ASX_HYDROXYL; 4.
DR PROSITE; PS00022; EGF_1; 5.
DR PROSITE; PS01186; EGF_2; 9.
DR PROSITE; PS50026; EGF_3; 9.
DR PROSITE; PS01187; EGF_CA; 3.
DR PROSITE; PS01209; LDLRA_1; 27.
DR PROSITE; PS50068; LDLRA_2; 32.
DR Calcium-binding; EGF-like domain; Endocytosis; Glycoprotein; Receptor;
KW Repeat; Signal; Transmembrane.
FT SIGNAL 1 20 Potential.
FT CHAIN 21 4599 Low-density lipoprotein receptor-related
FT protein 1a.
FT DOMAIN 21 4444 Extracellular (Potential).
FT TRANSMEM 4445 4467 Potential.
FT DOMAIN 4468 4599 Cytoplasmic (Potential).
FT DOMAIN 31 70 LDL-receptor class A 1.
FT DOMAIN 76 114 LDL-receptor class A 2.
FT DOMAIN 120 154 EGF-like 1.
FT DOMAIN 195 194 EGF-like 2, calcium-binding (Potential).
FT DOMAIN 255 336 LDL-receptor class B 1.
FT DOMAIN 338 380 LDL-receptor class B 2.
FT DOMAIN 382 424 LDL-receptor class B 3.
FT DOMAIN 568 609 LDL-receptor class B 4.
FT DOMAIN 611 655 LDL-receptor class B 5.
FT DOMAIN 657 705 LDL-receptor class B 6.
FT DOMAIN 707 754 LDL-receptor class B 7.
FT DOMAIN 794 834 EGF-like 3.
FT DOMAIN 844 882 LDL-receptor class A 3.
FT DOMAIN 885 923 LDL-receptor class A 4.
FT DOMAIN 926 963 LDL-receptor class A 5.
FT DOMAIN 966 1003 LDL-receptor class A 6.
FT DOMAIN 1005 1043 LDL-receptor class A 7.
```





Db 3056 FIYIDSSRPNGSRINRMCLNGSDIKVHVHTAPVNALAVDWIGKNLYWSDTEKRIIEVSK 3115  
QY 359 -----PT-----GNLL-ASC-----S- 368  
Db 3116 LNLGYPTVLVSKRLKFPRLSLDPAGNLYWIDCEYPHIGRVGMDGTNQSVIETKISR 3175  
QY 369 -----D-----M-----TL-K-I-W----- 376  
Db 3176 PMALIDYVNHLYWADENHIEFSNWDGSHRHKVPNQDIPGVIALTLFEDYIYWDGKTK 3235  
QY 377 --S-MK-----Q-D-----N-----C-----V 384  
Db 3236 SLRSHVKTSGADRLSLNLSWHAITDIQVYHSVRQPDVSKHLCTVWNGGSHLCLLGPQKT 3295  
QY 385 H-----D-----L----- 387  
Db 3296 HTCACPTNYLAADRNTCLSNCTASQFRCKTDKIPFWMKDVTVDGCGDSDEPDDCPEF 3355  
QY 388 -----Q-----A-----H-----NKE-I----- 394  
Db 3356 KCQPRFOCGTGLCALPAFICDGENDCGNSDELNCDTHVCLAGQFKTKNKKCIPVNL 3415  
QY 395 -----Y-----T-----I-K-W-----SP-----T 402  
Db 3416 CNGQDDCDEDEKDCPENSCSPDYFOCKTKTKHCISKLWVDEDPDCADASDEANCDKKT 3475  
QY 403 -GP-----G-----T-N-----N-P----- 409  
Db 3476 CGPHEFOCKNNNCIPDHWRCNDQNSDNDCKFQCTLKDPLCSNGDCVSSRFWCD 3535  
QY 410 -----N-----A-----N----- 412  
Db 3536 GEFDCADGDEKNCETSKDQFQCSNGQCLSAKWCKDGHEDCKYGEDEKNCPEAPFVCS 3595  
QY 413 ---LMLA-----SAS-----P----- 420  
Db 3596 SSEYMCASGGLSASLKNGBPCDVGSDMDCVIECKEDQFOCKNKAYCIPIRWLCDGI 3655  
QY 421 ---D-S---T-----V-R-----L-W-----D-V----- 429  
Db 3656 YDCVGSDEETCGRGSGICRDDEFLNNLSLKLHFWVCDGEDDCGDSDEAPDMCVKFLC 3715  
QY 430 -----DR-----GI----- 435  
Db 3716 PPTRPYRCNRDRI CLQLEKICNGINDCGDNDSEHSCGKLSLSPCKDDEFTCSNRNCI 3775  
QY 436 -----HT----- 437  
Db 3776 PMELQDSLDDCGGDSDEQGLKTPIEHTCENNGNPGDDAYCNQIKTSVFCRCXPFQR 3835  
QY 438 -----L-T-----K----- 440  
Db 3836 NMKGRECADLNECLLFGICSHHCLNTRGSKVCQDNQFQEKNNKCTAKGSEDOALYIAND 3895  
QY 441 -----HO-----E-----P-V-Y----- 446  
Db 3896 TDILGFVYPFNYSGHQOISHVEHNSRITGMVHYQRNVIVWSTQFPGGIFYKIMIDARE 3955  
QY 447 ---S-----VA-----PS----- 451  
Db 3956 KRQANSGLICPEPKRPRDIAVDWAGNVYVTDHSHRMHWFSSYYTHWTSLSRYSINVGQLNG 4015  
QY 452 -----P-----DG----- 454  
Db 4016 PNCTRLTNMAGEPYAIANPNKRGMMYWTVIGDHSHEEAAMDGTLRRVLVQKNLQRP TG 4075  
QY 455 -----R-Y-----LA---S---GS----- 461  
Db 4076 LTVDFGERIYWADFELSIIGSVLYDGSFPVSVSSKQGLLHPHRIDVPEDYIYGAPKN 4135  
QY 462 -----F-----DK-----CV 466  
Db 4136 GIFRVQKFGHGSVEVLALGVDTKTSILVSHRYKQLNLPNCPDLSCDFLCLLNPSGATCI 4195

QY 467 -----H-----I-----W-----N-T- 471  
Db 4196 CPBGKYMNGTCHDDSLDDSKLTCENGGRCILNEKGLDLRCHCWPSYSGRCVNHCSN 4255  
QY 472 --Q-----T-G-----AL-----V-----H-----S-----Y 480  
Db 4256 YCQNGGTCTPSTIGRPTCICALGFTGPNCGKAVCEDSCHNGGSCVVTAGNQPYCHQADY 4315  
QY 481 ---R-----G-----T-----GG----- 485  
Db 4316 TGBRCQYVCHYCVNSECTIGNDGSVEVCPTRYEGPKCEIDXCVRCHGHCIIKNON 4375  
QY 486 --IF-----E-----VC---W----- 491  
Db 4376 EDIFCNCNTNGKIASQCQCDGYCYNGGTQCLDPETSPVPCVCSNWSGTQCPERPAPKSSK 4435  
QY 492 -----C-----VLDLRK 514  
Db 4436 SEHISTRSTAIIVPLVLLVTLVTLVGLVVCCKRRTKTIRROPIINGINVEIGNPSY 4495  
QY 492 NA-----AG-----D-K---VG---A-----SA-----SD-----G-- 505  
Db 4496 NMYEVDHDSGDLLEPSFMIDPVKSRYIGGGSSAFKLPHTAPPIVYLSNLSLKGPLTFGPT 4555  
QY 506 --S--V-----C-----VLDLRK 514  
Db 4556 NYSNPVAKLYMDGQNCNRNSLASV-DERK 4583

Search completed: January 3, 2005, 15:26:03  
Job time : 135.333 secs